

Amber light for nuclear power

The long-awaited report on the future of nuclear power in Britain, a qualified licence to push ahead, is a proof that governments cannot delegate the big decisions.

PEOPLE who live in glass houses are often advised not to throw stones, for fear of the damage that might be done, but there is no reason why the injunction should apply to government inspectors such as Sir Frank Layfield, who for the past four years has been occupied with the single question of whether a pressurized-water reactor station should be built in Britain. The question is important because it was defined as such nearly ten years ago, when the previous British government told the Central Electricity Generating Board (CEGB) that it could have its wish, to build a novel reactor-type in Britain, only if there were a full-dress public inquiry in advance. By thus investing what should have been an engineering decision with great and artificial significance to both sides — those who would build nuclear power stations and their detractors — the then government and its successor, the present, made sure that the inspector would be fully aware of how high were the stakes. They put him in a glass house, and he has done his level best to see that the stones he has thrown, many of them effectively, are mostly little stones unlikely to shatter much glass. Nothing much else could have been expected in the circumstances.

The report is thus not the full-throated endorsement of expansion of nuclear energy in Britain for which some in the nuclear industry may have been hoping. Equally, there is no great comfort for the anti-nuclear lobby, although the Friends of the Earth have won a promise that CEGB should publish the final version of its safety plan when that is ready, so that interested parties can see whether their particular objections have been met. The most trenchant passages in the report are those in which the inspector takes elements of the bureaucracy to task for failing to talk each other's language, or simply to talk to each other at all. CEGB and the National Nuclear Installations Inspectorate plainly have to get to know each other better even after all these years. It would be wrong to say that Layfield sits on the fence, because there is nothing in the report to dissuade CEGB from going ahead, taking a few extra precautions along the way. But, ingeniously, the report will also make the fence broader at the top, on which it will be a little more comfortable to sit.

It may also seem strange, but is hardly surprising, that the inspector is able to complain that more than two years of evidence failed to answer all the questions that can be asked about the safety of pressurized water reactors; the complaint, of course, is directed at this method of public inquiry, which requires that all the arguments should be channelled through the head of a single person, the inspector. The Sizewell inquiry was made even more cumbersome by the requirement that all the evidence should be delivered orally as well as in writing, a way of making sure that the weight of evidence could not physically overwhelm the hapless man behind the desk, which is not a good likeness of how teams of professional engineers inquire into matters of safety and performance, which is even more labour-intensive. If there is a moral, it must be that governments cannot hope to delegate to inspectors their responsibilities for deciding the framework in which huge technically-complicated decisions should be made.

The irony of the Layfield inquiry is that it finished taking evidence before public anxiety about nuclear safety was raised

by the accident at Chernobyl last year, but that its recommendations will have to be turned into practical decisions by a government and public corporations such as CEGB all too conscious that the climate has changed. This, no doubt, is why the government has chosen to publish the report without saying what its own response will be. Rumours that the British government would pin its convictions to the mast of nuclear power may have been modified by the prospect of a general election within the next sixteen months. The difficulty now, given the rules of these inquiries, is that nothing can happen until the government gives some kind of answer. The most likely and perhaps the best outcome will be that the Department of Energy will rub CEGB's nose in the points on which Layfield has identified a need for change and then say that CEGB must make up its own mind.

That, Chernobyl or not, is what should have happened in the first place. In relation to nuclear power and other novel industries, governments have a proper role in the development of safety standards and the mechanisms for making sure that standards are enforced. Britain also has a parliament which can, in principle, ensure that the safety standards in force at any time are those that allow voters to sleep easily at nights. The British government, like some others, also has four nationalized electricity utilities (one in Northern Ireland, two in Scotland and one for the rest of Britain) whose needs of capital are potentially a strain on the public sector borrowing requirement; but that is an almost academic question when the issue is that of whether a single nuclear power station should be built by a cash-rich industry. There is also a Department of Energy, which oversees the utilities and the nuclear industry, which has, among other things, a responsibility for spending money on research directed towards the production of energy of all kinds. Constitutionally, in other words, there are many ways in which the issues delegated to Layfield could and should have been dealt with by the agencies and institutions to which the responsibility constitutionally belongs. If there were all the time in the world, the lucid Layfield report would be a good starting-point for many of these decisions. But the waiting time has long since been used up. □

The price of secrecy

The British government's latest self-made secrecy issue exemplifies its over-zealous policies.

THE British government seems perpetually in a jam about the secrecy of its own most secret activities, espionage and counter-espionage. For days on end last year, Britain's most senior civil servant, Sir Robert Armstrong, stood in the witness-box of an Australian court-room giving reasons why a former employee of British counter-espionage (called MI5) should not be allowed to publish some tales from his working life; it will be several weeks before the judge decides whether the British case is irreparably weakened by the government's apparent indifference two years ago to the publication of some of the same tales in a book by a British journalist, apparently with the connivance of the intelligence services and even, improbably, Lord Rothschild (see *Nature* 324, 396; 1986). Now, as if to make up for slackness then,



the government is going to extraordinary lengths to suppress news of a plan to build a reconnaissance satellite for recovering information from other people's internal radio transmissions. This coincides with suggestions that the intelligence services have been eavesdropping on the Irish government's private talks about Ulster, but it is unlikely that even the British would deem a satellite necessary for spying on Dublin.

The circumstances are these. A week ago, it emerged that the British Broadcasting Corporation (BBC) had decided, after several weeks of agonizing, not to broadcast a television programme that would have told the tale of the espionage satellite, said to be called Zircon. It emerged that the government had 'advised' the BBC that publication would be unwelcome. The author of the programme, Mr Duncan Campbell, a journalist whose strong suit is itself a kind of intelligence work — the piecing together of interesting tales from published sources, telephone directories, for example — promptly arranged to publish the essence of his story in the political weekly the *New Statesman*. There followed a series of pantomimic happenings — a government attempt to secure an injunction against Campbell (which failed because British judges can restrain only publications), an attempt by the Speaker of the House of Commons to prevent Members of Parliament from seeing the withheld film privately and a police search of the office of the *New Statesman* and of Campbell's own house. The essence of Campbell's charge is that the British government had been bent on concealing the project not only from the world at large but from the British Parliament. The Prime Minister, Mrs Margaret Thatcher, says that people will be "disgusted" that secrets have been irresponsibly revealed. The leader of the largest opposition party, Mr Neil Kinnock, appears to agree, but accuses the government of incompetence in the management of secrecy.

Several issues arise, of which the chief is the futility of attempting to conceal projects such as that Zircon is said to be. One of Campbell's sources of information appears to have been a series of discrepancies between the British government's published plans for building military communications satellites, many of them proudly announced by the companies awarded the contracts to build them, and the facilities on the ground for handling the signals they would generate. If a mere journalist can infer as much, are not other people's intelligence services likely to know what is afoot? And what purpose can concealment serve in the early stages of such a project when it must be plain that the cat will be out of the bag once the satellite is in place in geosynchronous orbit, where an antenna capable of collecting ground signals could not fail to be a conspicuous radar target? So much is openly recognized in, for example, the United States where, now that the convention that people do not seek to destroy others' satellites has apparently been accepted, the fact of surveillance satellites is openly acknowledged. Secrecy centres instead on what matters — the sensitivity or the resolution of particular devices. It may be inconvenient to be giving other people a few months' advance notice of what is intended (lengthened in this case by the hiatus in US shuttle flights), but what else is to be done in a democracy?

The British government's secrecy about Zircon cannot but reinforce the suspicion that its real anxiety is that otherwise it might be necessary to argue the case for spending \$500 million, by one estimate the amount needed over five years to put British academic research back on track (see p.377), on this project rather than on others. It is not, after all, that Zircon would have been the first of its kind; US satellites (called Signet) have been in the sky for some years, as have prototypes of the infrared sensors widely canvassed as the work-horses of the Strategic Defense Initiative. Two lessons stand out. First, as the Australian case shows, the most closely guarded secrets eventually become public knowledge. Second, democracies must calculate that the relatively rapid erosion of secrets is the part of the price they pay for their other benefits, among which that of a free press is too lightly scorned in Britain. □

Unquiet Chinese science

Chinese science is being blamed for the unrest among students. Will reform now be halted?

THE news from China is not good. Last week it emerged that Professor Fang Li Zhi had been dismissed as vice-president of the Hefei University of Science and Technology (see *Nature* 325, 290; 1987). Now the president of the Chinese Academy of Sciences, Lu Jiaxi, and vice-president, Yan Dongsheng, have been replaced, apparently at the instance of the National People's Congress, the nearest thing in China to a parliamentary assembly. It is as if Dr Frank Press were removed as president of the US National Academy of Sciences by a resolution of Congress, to be replaced by someone chosen by the White House.

The most cheerful reading of events is that the academy is merely being held responsible for having allowed Fang to become an outspoken advocate of democracy in China while playing a key part in the administration of the academy's model university, a remarkable institution that migrated from Beijing to Anhui province to escape the rigours of the Cultural Revolution, and which has ever since enjoyed a measure of independence that would not have been feasible in the capital. In that case, the recent dismissals may be the end of the Chinese government's reaction to last month's student demonstrations in Shanghai, Beijing itself but also in Hefei. The darker reading, and the general fear, is that these sackings are the beginning of a wider repression of free-thinking intellectuals.

For the past 15 years, the government of China has followed an enlightened policy in the encouragement of external links between Chinese researchers and those elsewhere. Both the international community and China itself have benefited enormously from this openness. It is inevitable that some of those who have been personally involved in these exchanges should have come to recognize that there are other and better ways of running research institutions and even governments than those with which China is lumbered.

Paradoxically, some in the Chinese government seem to have been looking for just such a falling of scales from clouded eyes as a by-product of the opening to the West. But both the government of China and the intellectual community are divided in their opinions of these developments. Who can quarrel with the opinion of China's hardliners that there are basic contradictions between the policy of openness towards the West and the maintenance of the bureaucracy of the world's largest and, many Chinese claim, most orthodox marxist state? How, for example, is it possible to reconcile the bourgeois liberal belief that researchers are best employed on what excites their interest with China's bureaucratic method of telling people where to work?

Wherever the upheaval of the past few weeks may lead, it is now inevitable that liberal opinion in China will be stifled, at least temporarily. It is ironical that, in a moving speech at Shanghai at the end of last year, Fang should have been congratulating himself and his colleagues that there was no longer reason to be afraid of speaking out in China. Events have proved him wrong, and the lesson will not be lost on his colleagues. The best hope now is that the damage will be limited. China has already won great benefits from its investment in research since the Cultural Revolution, and from the faltering steps that have been taken to weaken the political constraints on the interests and energies of researchers and to erode the bureaucratic impediments to accomplishment. There is no reason why these processes should not continue while the government of China seeks to come to terms with the underlying contradictions of its policies over the past 15 years. One danger is that it will mistake symptoms for causes, and put even internal reform on hold. The greater danger is that the Party Congress this autumn will take fright at the risks implicit in the contradictions, and turn back the clock to the ending of the Cultural Revolution — or even earlier. □

New science plans provoke UK government hostility

- BNSC's space plan too ambitious
- Value-for-money assessment needed

London

THE British government is seriously questioning the principal findings of two major reviews in science which could mean a rejection of their recommendations. Both the 15-year space plan and the blueprint to strengthen the British information technology industry are under fire from the Cabinet Office, which considers the schemes in part too ambitious or irrelevant.

The space plan argues that Britain must immediately invest £200 million a year — twice the present budget — to meet existing international commitments, principally collaborations involving other members of the European Space Agency (ESA). The Cabinet Office, whose scientific advisers seem to be an important influence in moulding government policy in science, now favours a more phased all been allocated.



Roy Gibson, head of BNSC, wonders how much he will be able to spend this year.

approach to the space plan under which the recommended yearly expenditure will be reached within three years.

Officials at the British National Space Centre (BNSC), the authors of the space plan, have become embarrassed as the government has failed to respond to the recommendations, delivered last summer. The embarrassment has become acute in the past two months as the officials, now meeting their European colleagues every week, are unable to commit Britain to any new project. Rejection of the plan will mean that BNSC must devise another, although it is becoming clear that little new money, if any, will be available.

A similar fate seems to have befallen the Bide report, prepared by a 21-strong team of industrialists and academics, led by the ex-chairman of Glaxo, Sir Austin Bide, and submitted to the government last November. The committee (called the IT86 committee) devised a strategy to continue the work of the Alvey research programmes, whose £350 million (about £200 million from government coffers) has

The committee recommended that the government should spend £500 million on new research programmes that would act as catalysts in attracting the same amount from industry. The Cabinet Office is unhappy about the report and seems to want some evaluation on previous Alvey research projects before any new financial commitment is made by government.

But the rejection of these plans, at least in part, will be seen by many as the latest examples of the government's reluctance to spend money on any significant scale.

The government's intransigence was challenged last week by the campaigning society Save British Science (SBS), whose membership is composed of about 700 prominent British scientists. According to the society, at least those research proposals put in the alpha class by the research councils must be supported, if only to restore the confidence of researchers. That would require an investment of some £100 million a year "for the immediate needs of top quality research".

In a letter to Mr Kenneth Baker, Secretary of State for Education and Science, Professor Joseph F. Lamb, chairman of SBS, emphasized the need. He wrote that "to restore, confidence the government must also announce a long-term policy of increased investment in civil science and technology. The case is urgent, we call upon the government to act without delay."

One report, however, has found favour with the Cabinet Office, that produced by the House of Lords Select Committee on Science and Technology and published on 8 January. While concluding that "neither government nor industry is spending enough at present levels to restore the UK's industrial position and markets", it did not recommend a particular figure for new investment in science. One of its principal recommendations is the creation of a minister at cabinet level, to coordinate research and development policy, and a Council on Science and Technology, to be chaired by the Prime Minister.

The idea of a council is well regarded in the Cabinet Office, which is keen on creating such a body, with members culled from industry, academic institutions, the financial world and government, to devise priorities in research and to ensure that current spending is cost-effective. The government is expected to respond to the Lords report by the spring.

Bill Johnstone

Layfield gives approval to Sizewell B

London

BRITAIN'S proposed £1,300 million pressurized water reactor (PWR) to be based at Sizewell in Suffolk and planned to be the forerunner of five similar complexes throughout the United Kingdom, has been recommended in the report* of the official inquiry, despite unanswered questions on safety. The report, prepared by Sir Frank Layfield who headed the study, was published this week and details the evidence submitted during the 340 days of the inquiry which closed in the spring of 1985 and in total has cost more than £3.7 million.

While the report firmly concludes that "on economic grounds alone, the project should go ahead", it continues: "The examination of safety was not exhaustive. The most important safety problems, and those which cause particular public anxiety, were examined in depth. But many matters were examined only briefly if at all, and much work remains to be done on aspects of the safety case ... Furthermore, the design was incomplete and not yet susceptible to a final assessment of its safety".

For those reasons, claimed Layfield, he "must place a great deal of reliance" on the evidence and promises of the Central Electricity Generating Board (CEGB), which will operate the reactor, and the Nuclear Industry Inspectorate (NII), the safety watchdog of the industry. Layfield claims that they can provide a "continuing assurance of safety". This conclusion is likely to provoke criticism from those who submitted evidence to the inquiry questioning claims made by CEGB and NII.

Since completion of the inquiry, the accident at Chernobyl has heightened public awareness of the dangers. An accident at Sizewell B would have "tolerable" consequences, says Layfield. "Theoretically possible accidents which could cause hundreds or thousands of deaths would almost certainly not occur."

Layfield remains confident that within his best estimate only "one or two workers" at the reactor would die of radiation-induced cancers and that there would be no radiation-induced hereditary diseases.

The economic case for the reactor is fully supported by the Layfield study, as it dismisses the case for an alternative to nuclear energy, principally coal. "The probability of a coal station having lower costs is remote", concludes Layfield. Bill Johnstone

*See page 378 for more on Sizewell. The 'Summary of Conclusions and Recommendations' of the Sizewell B Public Inquiry costs £4.95 from HMSO. The whole report, in 8 volumes, costs £30.

Sizewell inquiry leaves questions unanswered

London

DESPITE having placed much confidence in the assurances of the Central Electricity Generating Board (CEGB) and the Nuclear Installations Inspectorate (NII) on the safety aspects of the pressurized water reactor (PWR), the Layfield Inquiry was scathing in its criticisms of the poor communications between the two bodies.

Managerial and procedural weaknesses stand out despite the complexity of the licensing processes, which require "highly efficient exchange" of complex information. Layfield claims that although no evidence was submitted that suggested that those managerial weaknesses "have affected the thoroughness with which safety problems have been tackled or the level of safety achieved by the proposed Sizewell B design", there was stark evidence of confusion over responsibility and agreed safety standards.

Neither NII nor CEGB was totally responsible for these ambiguities, claims the report, and the government must share the blame. There has been little or no parliamentary or governmental guidance on how the licensing process should be carried out or on the basis for safety assessment, says Layfield.

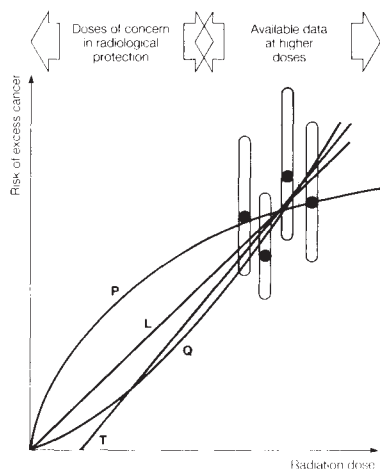


Fig. 1 The black circles show the excess of cancer observed in four groups of people exposed to different absorbed doses of X-rays. The vertical error bars represent the range of uncertainty in the observations. Four assumptions were suggested for extrapolating these data to obtain a relationship between dose and rate of excess cancers at low doses. (1) The excess of cancer is proportional to the absorbed dose, as represented by the straight line. (2) A relationship represented by the curve Q, whose general shape is developed from experimental and theoretical radiobiology, but whose exact shape cannot be predicted for humans. (3) A relationship represented by the curve P, which would fit excess cancers observed in workers at the Hanford atomic plant in the United States. (4) There is a threshold dose below which there is no risk of radiation-induced cancer, as represented by the line T.

The high level of mutual understanding needed between CEGB and NII is not helped by the "lack of sufficiently clear and agreed safety criteria" and the existence of different safety criteria that "was perplexing to the objectors and is likely to be a source of confusion to the layman".

One of the most serious deficiencies in the relationship between the two groups was weak communications. This led to frequent misunderstandings and on occasions to "fundamental misapprehension" of each other's position.

Layfield concluded that the submission of evidence during the inquiry showed the serious weaknesses in the relationship between NII and CEGB. During the course of the inquiry, "important initiatives" were taken to improve the management but "it was too early to judge the efficiency of these new arrangements".

The susceptibility to radiation and its effects are the two primary issues that originally precipitated the inquiry. The report details much of the known effects of radiation. The principal effects of low-level radiation, the report concedes, are to increase the probability of cancer or hereditary disease which may not occur for years. The short-term ill-effects "only occur in unusual situations such as a serious accident involving radiation". The primary public fear is the development of cancer as a result of either small or large exposures to radiation.

Because cancer is a common disease, now claiming one in every five deaths, it is difficult to assess easily the cancer that would not be expected to occur naturally, claims the study. Says Layfield: "In groups exposed to ionising radiation the excess of cancer due to radiation is likely to be small compared with variations in the natural incidence of cancer. Estimates of excess cancers are therefore likely to be subject to considerable uncertainty."

The exact shape of the dose-response relationship for radiation-induced cancers cannot yet be predicted "with confidence for humans". Four assumptions have been made about this relationship (Fig. 1).

The evaluation of the risk is as important as the risk itself. Nuclear power is widely considered, says Layfield, to be unique in the nature of the risks it poses but "risks similar to those often thought to be unique to nuclear power are associated with some other industries".

Carcinogenic chemicals that are imperceptible to the senses can be released in an accident in a chemical plant. And techniques for numerically estimating the risks from nuclear power have only recently been developed, says the report, and "the results of such estimates are subject to considerable uncertainty".

On the criteria available, Layfield has

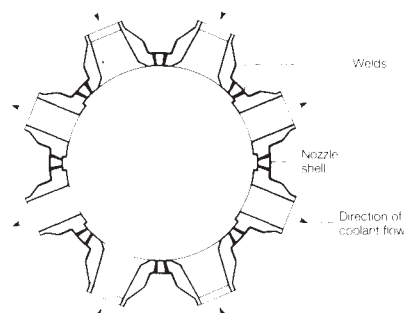


Fig. 2 The main differences between the conventional integral design and that to be used at Sizewell (above) are: (1) The shell flange and nozzle ring are made of a single forging in the integral design as opposed to two forgings welded together. (2) In the integral design the nozzles are not welded into holes in the nozzle ring, but to its outer surface. Thus the holes in the nozzle ring correspond to the inner rather than the outer diameter of the nozzles. (3) In the integral design the weight of the vessel and its contents is supported by special stubs attached to the outside of the vessel, whereas the Sizewell B design is supported on four of the eight nozzles. Thus the integral design potentially reduces stresses where the nozzles join the nozzle shell, a region where the stress distribution is complex. (4) The thickness of the integral vessel is greater than that of the Sizewell B vessel.

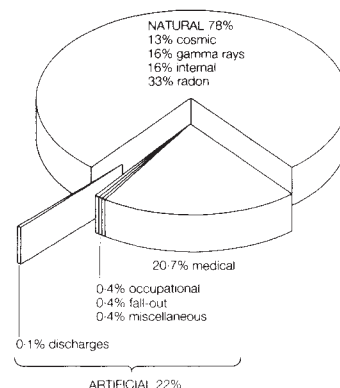


Fig. 3 The sources of radiation to the UK population.

determined what he considers a "tolerable risk". Any system of risk evaluation must take account of the special factors that might distinguish that risk, says the study. The nuclear industry is unique in the eyes of the public Layfield admits.

"The attitudes of members of the public who were worried about the safety of Sizewell B demonstrated that any system of risk evaluation must take account of special factors which distinguish the risk being examined. Many people regard the risks from nuclear power as having a special character, with the consequence that these risks should be regarded with more seriousness than other risks nominally of a similar size. One way of deciding on a tolerable level of risk is to find out what risks the public is already known to tolerate" (Fig. 3).

Bill Johnstone

Chinese and Japanese outshine US schoolchildren in maths

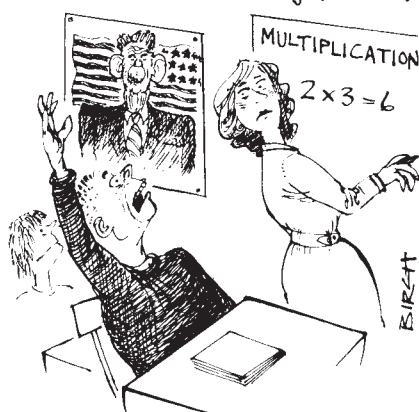
Washington

THE abysmal mathematical skills of US schoolchildren as compared with their Chinese and Japanese counterparts have prompted a new wave of concern about teaching methods and curricula in US schools. The newly formed Mathematical Sciences Education Board (MSEB) is so concerned it is developing the first national guidelines for mathematics curricula, a radical act in a country where the independence of local school boards is a cherished political freedom.

That typical US schoolchildren are easily outperformed by Chinese and Japanese children at all ages from kindergarten up has been known for some time. Possible reasons — and solutions — were discussed two weeks ago at a meeting held by the board in Washington, DC.

The figures are disturbing: according to a new analysis of the Second International Mathematics Study presented at the meeting, US students at the ages of 13 and 17

So how come the more missiles you sell, the less hostages you get back?



performed well below the international average of 17 countries in the study at problem-solving tasks (as distinct from computational tasks). In some, the US students were ranked in the lowest quarter of the countries sampled and the lowest of the advanced industrialized countries.

The reasons may not be hard to find. According to a study presented by James Stigler of the University of Chicago which compared US, Chinese and Japanese pupils, US children at the ages of 6 and 10 spend substantially less time engaged in "academic activities" than their Asian equivalents. By the age of 10, US children are spending fewer than 20 hours per week at work, as compared with 40 hours in China and 32 in Japan. Furthermore, the percentage of study time devoted to mathematics was lower in the United States than in the other two countries, and the children spent more time engaged in "in-

appropriate activities". Stigler comments that Chinese and Japanese teachers appear better prepared for teaching mathematics than US teachers.

Whatever the reasons, the United States' poor showing is deemed a matter

Multi-million dollar vaccination drive against infant mortality

New Delhi

IN the next five years, at a cost of US\$250 million, 82 million infants and 92 million expectant mothers will be immunized in India against six childhood diseases — polio, tetanus, diphtheria, measles, tuberculosis and whooping cough — in an ambitious programme aimed at reducing infant mortality by half before the end of the decade. At present, one in every nine babies dies before its first birthday.

The Universal Immunisation Programme (UIP) will involve administering 820 million doses of vaccine. It will be implemented as part of a primary health care programme through a network of 30,000 grassroot workers, each of whom will be given a motorized cycle and a vaccine carrier (ice box). The vaccine carrier is the last and the cheapest item in the \$60 million's worth of cooling equipment that includes refrigerated trucks, 275 walk-in coolers, 2,125 deep freezers, 28,000 ice-lined refrigerators and 800 vans for transporting vaccines.

The programme is backed by the World Health Organisation (WHO), UNICEF and other international agencies. But in spite of its laudable objective, UIP is seen by critics as one more programme imposed by outside agencies without regard to the primary health-care philosophy of self-help.

"By inhibiting community self-reliance and social control over medical technology . . . our people are once again being made dependent on Western countries for funds, vaccines and equipment", says Dr D. Banerjee, professor of social medicine at the Jawaharlal Nehru University in New Delhi.

India does not manufacture measles vaccine, which has to be imported together with much of the cold equipment, such as refrigerators running on sunlight or kerosene. To meet the target, India has to double its production of other vaccines or rely on imports. Plans are already under way to import technologies for making safer vaccines. According to WHO's own estimates, its immunization programme in the developing countries is turning into \$25-million-a-year business

of serious concern by officials in the White House's Office of Science and Technology Policy. Education Secretary William Bennett has shown an awareness of the problem by publishing his own study of Japanese education, although he supports the Reagan administration line that there should be no national curriculum like that in Japan. If MSEB has its way, there will at least be a standard against which curricula can be measured.

Tim Beardsley

for industrialized countries which now supply 85 per cent of the equipment and products for cold storage.

Concern has also been expressed about the wisdom of a massive vaccination programme in the absence of a mechanism for strict quality control of the vaccines. Ten children died recently in a Bombay hospital after receiving DPT (diphtheria/pertussis tetanus) vaccine made by one of India's premier laboratories. There have been several cases of polio among vaccinated children.

Banerjee says that "saturation spraying with vaccines" has been launched without basic epidemiological data. "The imposition of a selective approach, in the short term, in the name of saving children, is diverting scarce resources from the underlying need for comprehensive primary health care."

K.S. Jayaraman

When the wind blows



Jim Bloggs reads up on how to survive the holocaust.

THE effects of a nuclear attack have been animated in the form of a cartoon version of Raymond Briggs's successful book and radio/stage play *When The Wind Blows*. The film features the voices of Sir John Mills and Dame Peggy Ashcroft as Jim and Hilda Bloggs, the heroes of the film, who outline their fears and hopes as they find themselves survivors of a nuclear attack. *When The Wind Blows* can be seen around Britain from 6 February.

Almost naively, Jim and Hilda, a typical English retired couple, attempt to face the aftermath of a nuclear attack "with the same faith and optimism with which they once faced the Blitz". □

Kazakh Academy of Sciences in trouble with Moscow

London

THE Academy of Sciences of the Kazakh SSR has failed to carry out tasks vital to the Soviet economy, according to a representative of the Central Committee of the Communist Party of the Soviet Union. Mikhail Solomentsev, chairman of the Party Control Committee of the Central Committee, visited the Kazakh capital, Alma-Ata, in the wake of the student demonstrations of 17–18 December, and castigated the Kazakh academicians on several scores. They had, he claimed, failed to resolve vital problems of the integration of science and production in due time; there had been “false accounting” and “infringement of the principles of selection and education of cadres”, and there had been a lack of the *glasnost* (openness) that is the official keynote of the Gorbachev regime. The academy, Solomentsev concluded, exhibits a “substantial alienation” from the task of Kazakhstan’s “socio-economic development”.

Clearly the unrest had a significant nationalist content. The protests were sparked by the replacement of the retiring Kazakh First Secretary of the Communist Party of Kazakhstan by a Russian, but this in itself would probably not have been a sufficient cause. (As a Hungarian newspaper pointed out, the First Secretary of Kazakhstan in 1955–56 was Leonid Brezhnev, and he encountered no nationalist upheavals.) But during the past few years, there has been growing concern among the traditionally Muslim peoples of Soviet Central Asia that their interests have been subordinated to those of Russia proper and the Western republics of the USSR. To accuse the Kazakh scientists of neglecting the interests of their republic is thus to throw doubt on all Kazakh national aspirations. Solomentsev’s allegation that the academy had given less help than it is capable of in drafting environmental conservation measures must have been particularly stinging.

The students directly involved in the demonstrations have come in for heavy criticism in the media, particularly at the All-Union level. *Sotsialisticheskaya Industriya* claimed that some of them were “stimulated” by drugs. The young Communists’ paper *Komomolskaya Pravda* criticized the facts that the “overwhelming majority” of students in Kazakhstan were Kazakhs “although people of over 100 nations and nationalities live in the republic”, that they were housed in hostels according to nationality, and that the majority of trouble-makers came from the ethnically homogeneous southern regions of the republic, had attended schools

where Kazakh was the language of instruction and had only a poor knowledge of Russian.

Accordingly, top party officials (including the military commander of Soviet Central Asia) have been sent to the various universities and higher education institutions to address the students and staff on their socialist obligations, and a meeting was held last week between the new First Secretary of Kazakhstan, Gennadii Kolbin, and university and college rectors. Kolbin stressed a number of negative factors in the university admission procedures, “gross miscalculations” in the orga-

nization of education and training, and a “mistaken” cadre policy which had a negative effect on the students’ *weltanschauung*, which was, he alleged a contributory factor to the past months’ “hooliganism”. Meanwhile, urgent measures have been rushed through the Kazakh Council of Ministers to release funds for the construction of new student hostels by mothballing laboratory, computer and university administrative blocks already under construction.

Vera Rich

Other practical steps have been taken. The Young Communists’ secretary for the first-year physics courses at the Kazakh State University has recently received a seven-year sentence for “inciting” the December disturbances, and a student is now facing trial for perjury in connection with the case.

US–Japanese research initiative renegotiates cooperative projects

Washington

A HIGH-LEVEL delegation, led by White House Science Advisor William Graham, will travel to Tokyo next week to prepare for renewing an agreement with Japan, due to expire at the end of April, for the promotion of cooperative research and development in science and technology. One aim of the visit is to discuss what the United States considers to be imbalances in the treaty; at least one member of the delegation believes those imbalances need serious attention.

Frances Li, a senior policy analyst for the White House Office of Science and Technology Policy, says the United States would like to see a presidential agreement such as that for science and technology focusing on a few high-profile projects. Although she will not speculate on what those projects might be, she feels they must be of sufficient stature and importance to warrant the attention the bilateral agreement is receiving. The US delegation will also seek to spell out more thoroughly the basis for joint cooperation at all levels of research.

Government agencies included in the initial agreement, which was signed in 1979, have had mixed experience of the cooperative projects. At the National Aeronautics and Space Administration and the Department of Health and Human Services, joint programmes started immediately, and some flourished. But at the Environmental Protection Agency (EPA), none of the projects initially proposed got off the ground. Although EPA has 14 joint exchanges with Japan for sharing environmental information, none of these involves research cooperation and they are covered by a separate agreement.

John Moore, the National Science Foundation’s deputy director, will be a

member of the US delegation, together with assistant secretary John Negroponte from the State Department and assistant secretary Bruce Merrifield from the Commerce Department. Merrifield is one of those who believes that information has been flowing one way from the United States to Japan. He says that semantic difficulties have sometimes clouded issues of what is permitted under the agreement. But, says Merrifield, “in many cases, the Japanese have clearly understood what they’re doing and have taken advantage of us.” By erecting barriers to innovation, such as inappropriate government regulation and antiquated anti-trust laws, US research has been an easy target for exploitation, according to Merrifield.

Merrifield believes that the new Federal Technology Transfer Act signed into law last year (see *Nature* 323, 659; 1986) will change things. The law is designed to encourage federally supported research laboratories to seek commercial partners for in-house research. Merrifield believes this will cause the supply of ‘free’ information to dry up, which is why, he says, Japan is “desperately anxious to have this [agreement] renegotiated”. But he insists that will not happen unless the Japanese come closer to a more equitable sharing of resources and talent.

Japan appears sensitive to Western concern about cooperative research. One Japanese initiative, the Human Frontiers Science Program, would promote ‘internationalization’ of scientific research. Projected to cost 1 million million yen over its 20-year life, Japan has proposed a 200 million yen starting budget for 1987. But the Human Frontiers Science Program is more of a multinational effort, and may not be part of the bilateral discussions next week.

Joseph Pa’ca

Soviet academician makes rare appearance at US Congress

Washington

IN one of the extremely rare appearances of Soviet citizens before the US Congress, Academician Evgeny Velikhov, vice-president of the Soviet Academy of Sciences, spoke out on Chernobyl and the Strategic Defense Initiative (SDI). He was giving evidence to the Senate Labor and Human Resources Committee, chaired by Senator Edward Kennedy (Democrat, Massachusetts).

Velikhov fielded questions about the causes and consequences of the accident last April at the Number 4 reactor at Chernobyl, where his involvement began just days after the event. Much of his testimony reiterated well-known facts about Chernobyl. He blamed the accident on "multiple errors and deviation from procedure" that allowed the reactor core to overheat. The six operator errors leading to the accident were "unthinkable", according to Velikhov, and yet they happened. He specifically blamed the technicians involved for not discussing their plans with plant physicists and safety personnel.

Decontaminating the reactor site has been difficult. Although, Velikhov said, physical damage to the reactor building itself was surprisingly small, the radioactive debris deposited both near the reactor and in an area extending 30 km from the reactor was the real problem. Remotely controlled equipment was used to remove radioactive debris, but failures were common. Velikhov suggested that the United States and Soviet Union should collaborate on the technology of machines to operate in radioactive environments.

Thirty-one people died from injuries acquired at Chernobyl, although there was only one death from the fire and explosions on the day of the accident. Velikhov said that most of the other 237 who were hospitalized have now returned to work. In testimony later in the hearing, Robert Gale, the physician at the University of California, Los Angeles, who performed bone marrow transplants on victims of the accident, said that there may be as many as 150,000 additional cancers over the next 50 years attributable to radiation from Chernobyl, but he added that this number must be compared with the expectation that total cancers worldwide in the same period would number 1,000 million. Gale also said the Soviet Union is establishing a register of all 135,000 people evacuated from the Chernobyl area to study health effects of the accident.

The scope of the clean-up is staggering. Three hundred thousand cubic metres of concrete — enough to build a 60-m-high building — were used to entomb the dam-

aged reactor. Workers have also built a concrete wall 15 m deep and approximately 6 km long around all four Chernobyl reactors to prevent seepage of radioactive material into local rivers. Decontamination teams cleaned some 60,000 buildings in nearly 500 villages in the area around Chernobyl. Remarkably, a small number of people have already returned to the villages around the reactor.

Velikhov told senators last week that there are important non-technical lessons to be learned from the accident. It is easy, he said, to become lulled into a false sense of confidence after years of accident-free operation. He criticized the United States



Velikhov faced a Senate committee in a two-hour-long session.

for not making a stronger commitment to his own area of expertise, thermonuclear fusion research. But he conceded that Soviet dependence on nuclear power will grow for the rest of this decade. Plans now call for a fivefold increase in nuclear capacity by the year 2000. Two of the four reactors at Chernobyl are now back in operation, one running at 50 per cent power, the other at 90 per cent.

On SDI, Velikhov said his government had made a major concession at Reykjavik by agreeing to permit continued laboratory research on SDI. Velikhov stressed that, by "laboratory testing", he did not mean only work carried out inside a building; field testing would be allowed, but only with weapons designed not to perform at full operational power.

Speaking to reporters after the hearing, Velikhov pointed out that if supposedly fail-safe systems in a nuclear power plant could fail, surely the chance of error in a programme as complex as SDI could not be ruled out. Velikhov believes it is vital to limit offensive weapons, as the next generation of weapons will be extremely complex and more prone to accidental use.

While in Washington, Velikhov also met Frank Press, president of the National Academy of Sciences, to discuss US-Soviet scientific exchanges. He also talked with US organizers of an international meeting on arms control to be held in Moscow on 14-16 February. **Joseph Palca**

Companies to pay for using others' ideas?

Rehovot

COMPANIES that benefit financially from unpatented ideas used in manufacturing products should allocate 0.5 per cent of their earnings to fund the research of the scientists whose work they have exploited.

This firmly held view is that of Weizmann Institute Professor Meir Wilchek who has been a victim of patent law. The results of his research on peptide chemistry, affinity chromatography and the avidin-biotin system — like those of Nobel laureate César Milstein on monoclonal antibodies and of Professor Robert Gallo on interleukin 2 — were never properly patented.

Last week, together with Professor Pedro Cuatrecasas of Glaxo Inc. in the United States, Wilchek was awarded the \$100,000 Wolf Prize in Medicine for "the investigation and development of affinity chromatography and its application to biomedical sciences". Yet the fruits of Wilchek's work at the institute and at the US National Institutes of Health, which have helped commercial companies around the world earn tens of millions of dollars, have made only the slightest contribution to his personal finances and that of his research.

Wilchek is satisfied to live on the very modest \$1,200 a month in take-home pay he receives from the Weizmann Institute, but he feels that it is unfair, even immoral, that commercial companies should be the sole beneficiaries of someone else's efforts, even if that is their legal right. He therefore proposes that companies manufacturing products directly based on unpatented scientific studies should voluntarily set aside 0.5 per cent of their net earnings for research grants to scientists from whom they have 'borrowed' ideas. This would be far less than the sum paid to scientists with valid patents, but it would be enough to encourage unconventional research.

Wilchek welcomes the grace period now incorporated in US law, which gives scientists up to a year after publication of papers to patent their new compounds and processes. But he thinks that is not long enough.

"After all", he points out, "three years passed between the initial papers (published in 1972) which prefaced the development of the avidin-biotin system, and my own realization of its great commercial significance. By then, of course, it was too late to obtain a patent. And only in the early 1980s did the avidin-biotin system become big business."

Nechemia Meyers

Woman's AIDS death in Japan produces shock waves

Tokyo

THE first Japanese woman to be officially recognized as a victim of AIDS (acquired immune deficiency syndrome), died last week, awakening Japan to the fact that the disease is not confined to homosexuals and haemophiliacs.

According to the Health and Welfare Ministry, the number of AIDS cases in Japan totals 26; 18 patients have already died. There may be more unreported cases.

More than 50 per cent of Japanese AIDS cases are haemophiliacs; nearly all the others are homosexuals. The large percentage of haemophiliacs has arisen because of a prime infection route via imported blood products, mainly from the United States (see *Nature* 315, 8; 1986).

Blood tests since 1984 establish that about 30 per cent of Japan's 4,000–5,000 haemophiliacs have been exposed to AIDS. Japan's first AIDS victim was a haemophiliac who developed the disease in 1981, and antibodies to the AIDS virus have been detected in sera of haemophiliac patients dating back to 1980. It has been suspected since 1985 that thousands of homosexuals carry the virus, and the first homosexual victim developed AIDS in 1982. Yet only 14 haemophiliacs and 10 homosexuals have been officially recognized to have AIDS, fewer than 1 per cent of the inferred number of carriers.

There have been problems recognizing the disease. The latest victim, a Kobe prostitute, first approached a hospital in July last year complaining of fatigue and severe diarrhoea. In October she was diagnosed as having pneumonia and was transferred to a second hospital. Unresponsive to treatment, she was moved to a third hospital in December. Only then were blood samples taken to check for AIDS. She was officially recognized as an AIDS sufferer by the Health and Welfare Ministry on 17 January, three days before she died.

Doctors may be reluctant to report patients to the ministry. Those treating haemophiliacs are particularly concerned that their patients should not be ostracized by society or barred from schools. There is no legal requirement to report cases, only "administrative guidance"; patients with AIDS-related complex are not reported and nor are infection routes.

The ministry is considering the introduction of a new law to make AIDS a notifiable disease. But some members of the ministry's "AIDS Countermeasures Experts Conference", would like to ensure anonymity for haemophiliacs.

A third brake on the reporting of AIDS may be the ministry itself. Japan's first

AIDS victim died in July 1983. But the first report of AIDS was not released until March 1985, after one of the ministry's advisers had published the results of blood tests in a medical journal (see *Nature* 315, 8; 1985).

A chief concern is to avoid "panic and confusion" by setting up advisory services before making public announcements. In step with this, since November last year, the Japanese Red Cross has been screening all blood donors for AIDS and ATL (adult T-cell leukaemia) (See *Nature* 323, 384; 1986). But the ministry decided that although AIDS carriers would be told that they were infected, ATL carriers would not.

The reason is one of numbers and facilities. Only ten AIDS carriers have so far been found among two million blood donors, but more than 1 per cent of Japan's nearly 9 million donors are believed to carry ATL. Although only about one in 2,000 ATL carriers develop the disease, usually in their forties or fifties, ATL is fatal. As evidence mounts that the virus is transmitted from mother to child through mother's milk simple advice not to breast-feed could save lives. But the

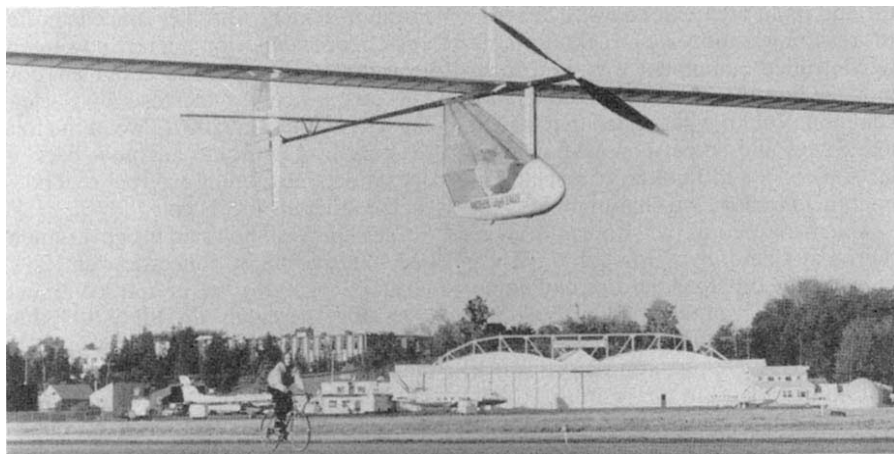
ministry is not notifying ATL carriers because of a "lack of appropriate administrative provisions".

Following the announcement of Japan's first female AIDS victim, there was a flood of calls in the Kobe area seeking advice and blood tests, mainly from men who had visited 'soapland' (massage parlours) and other prostitutes. The public is now aware that Japanese women can catch AIDS, but ignorance about the disease is still rife. At a public hearing on AIDS in Yokosuka last week, audience attention focused on such questions as whether AIDS can be caught from train straps.

The government has boosted funding for AIDS research in fiscal year 1987 to Y155 million (about US \$1 million), Y112 million up on last year. Part of the funds will be used to develop drugs to combat the disease, including a Chinese herbal medicine derived from liquorice. But the funds allotted to AIDS research and prevention in Japan pale into insignificance when compared with the over \$500 million requested for the 1988 US budget, and only Y7 million has been put aside for administration costs and public information. Leaflets on AIDS are available on request at local ward offices and health centres, but there have been no major publicity campaigns as seen in the West.

David Swinbanks

World human-powered flight record



Washington

A WORLD distance record for human-powered flight was set last week by Glenn Tremml, piloting a craft designed by Massachusetts Institute of Technology engineers. Tremml stayed aloft 2 hours, 13 minutes and 14 seconds, travelling 59.9 km (37.2 miles) over the Mojave desert at Edwards Air Force Base in California. Earlier last week, pilot Lois McCallin set a woman's distance record with a ten-mile flight. The craft, the Michelob Light Eagle, is a prototype for a second plane that will attempt to fly 111 km (69 miles) from Crete to the Greek mainland later

this year, recreating the feat of the mythical Greek character Daedalus. The plane flies between 1.5 and 3 metres (5–10 feet) off the ground and averages about 25 km (16 miles) per hour. It weighs 40 kg (88 lb), is 9 metres (29.6 feet) long, and has a wingspan of 31.1 metres (102 feet). The aircraft frame is constructed of a carbon-fibre material called Thorneil bonded by epoxy into a strong, lightweight design. One hundred and twenty ribs of foam and aircraft plywood give the plane aerodynamic shape. The wings are covered with plastic, and the body of the plane with a synthetic cloth.

Joseph Palca

Shuttle parts could provide route to cut-price heavy launcher

Washington

THE announcement that the Strategic Defense Initiative (SDI) is seeking \$110 million to develop technology for a new heavy-lift launch vehicle has prompted a surge of interest in plans for future satellite launchers. Although there are still no firm plans for a launch system with capabilities superior to the shuttle, the backlog of launches left by the Challenger accident, together with the huge demand expected should a ballistic missile defence system be deployed, is creating pressure for an early decision. Chief among the requirements are increased payload weight and reduced launch costs without sacrificing reliability.

The SDI request is part of a \$2,800 million supplement sought for the current fiscal year defence budget. But the National Aeronautics and Space Administration (NASA) has also been making noises about a future heavy-lift vehicle, and officials there hope that agreement will be reached with the Department of Defense (DoD) jointly to develop an unmanned launcher. It will be only partly reusable and therefore may not meet all SDI requirements.

SDI critics such as the Federation of American Scientists claim that the heavy-lift technology research effort pre-empts a decision on whether to deploy a missile defence system. DoD replies that it is prudent to have a heavy-launch capability should one become necessary.

In the short term, there are several proposals for a launch vehicle using components of the shuttle system. Such a launcher could be ready in the early 1990s, and could use existing launch facilities in Florida and California. NASA believes a shuttle-derived vehicle could increase payload weight to low Earth orbit by a factor of three or four compared with the shuttle, bringing the total to 150,000–200,000 lb. The cost of a launch would be similar to that of a shuttle launch. A decision to build such a vehicle might, however, delay developing a more advanced fully reusable launcher that some believe will ultimately be necessary for SDI.

Among the shuttle-derived proposals offered to NASA is one from Boeing Aerospace and Hughes Aircraft that might be produced commercially. The 'Jarvis' craft would be capable of putting about 80,000 lb into low Earth orbit; on present plans, it would use two shuttle solid rocket boosters and one shuttle main engine. The payload would be mounted above the external tank used by the shuttle. United Technologies has proposed a design in which the payload would be side-mounted, replacing the shuttle

orbiter, on a stack consisting of an external tank and two solid rocket boosters. United says its concept could lift 115,000 lb to low Earth orbit; later versions could be made fully reusable. Martin Marietta also has several design concepts incorporating shuttle components.

To realize the improvements in launch costs and in flight rates considered necessary for SDI, many believe a separate and fully reusable system must be developed. Stanley Lewis of Rockwell International, one of four industrial contractors working

on the inter-agency Space Transportation Architecture Study, believes complex military satellites for SDI could cost up to \$4,500 million apiece. Although projections of heavy-lift requirements vary, Lewis believes a fully reusable unmanned launcher capable of lifting up to 150,000 lb into low Earth orbit makes sense for the late 1990s.

The Space Transportation Architecture Study, which examines the longer-term possibilities and requirements, will be completed in about a year. But SDI is not in a mood to wait, and Defense Secretary Weinberger is now said to want to move "briskly". One way or another, decisions on future launch capabilities are likely to be made before then. **Tim Beardsley**

India's plans for an indigenous launch vehicle still on target

Bangalore

INDIAN scientists are in the final stages of preparation for two major space projects — the launch of the second generation Augmented Satellite Launch Vehicle (ASLV) from India itself and the Indian Remote Sensing Satellite (IRS) from the Soviet Union.

At the same time, efforts are being made to keep the multipurpose domestic spacecraft INSAT-1C ready for launch by

vehicle. The launch of this rocket, an improved version of the first-generation SLV-3, was postponed from 1985 because of problems in the motor of the fourth stage.

During its second flight in 1988, ASLV will carry a joint Indian/West German payload for remote sensing with stereoscopic equipment. For the third ASLV flight, the Indian Space Research Organisation (ISRO), in conjunction with the Council for Scientific and Industrial Research (CSIR), will develop a satellite payload to monitor the upper atmosphere.

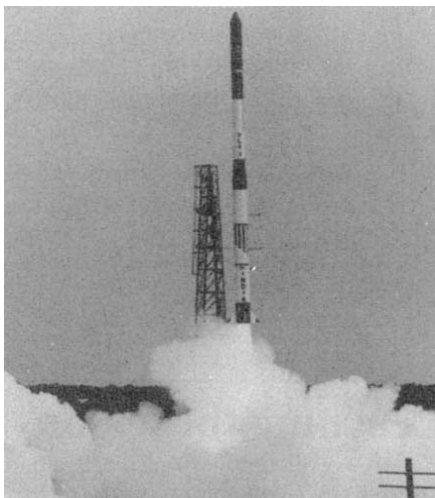
IRS, to be launched from the Soviet Union, will be lifted into a polar orbit by a vehicle provided by the Soviet Glavcosmos. The satellite will gather data on agriculture, forestry and minerals and will be the first satellite for which India will bear the launch costs. Three earlier satellites were launched free by the Soviet Union.

A major mapping project has been planned for this year covering 183 districts, for which India will depend on data from the US Landsat and the French Spot spacecraft.

Another crucial area in which Indian space engineers are working is in the recovery of rocket boosters. It is proposed to recover the first-stage booster of PSLV with parachutes. This will save India more than Rs400 million. Four launches of PSLV are planned by 1995. The first stage forms the core booster with 125 tonnes of solid fuel and an additional six strap-on boosters, each with 9 tonnes of solid propellant.

Mitsubishi Electricals of Japan declined some time ago to provide India with LE-5 cryogenic engine technology, but Professor U. R. Rao believes that India would not find it difficult to develop an indigenous launch vehicle to carry a geostationary satellite, based on cryogenic engines.

Radhakrishna Rao



India's first-generation launcher, SLV-3.

the European Ariane vehicle in February next year. India is also developing a bi-propellant Polar Satellite Launch Vehicle (PSLV) and a liquid fuel Geostationary Launch Vehicle (GSLV), both of which are expected to take off in the late 1980s or early 1990s.

The five-stage, solid-propellant-based ASLV, planned for launch in late March from India's eastern missile range on Sriharikota Island, is designed to put a 150-kg spacecraft into a near-Earth orbit. The satellite will carry a payload for the study of gamma rays as well as a package for monitoring the performance of the launch

Castles in the Saharan air

SIR—Many measurements have shown that the ozone layer in the troposphere is unstable, that carbon dioxide, carbon monoxide, methane, oxides of nitrogen, halocarbons and other trace gases and dust are increasing in the atmosphere, and that sea surface temperatures are warming, glaciers are receding and the sea level is rising¹⁻¹⁰. I suggest a scientific project for correcting the buildup of CO₂.

I propose a feasibility study of an attempt to establish new flora and fauna in the Sahara, the world's largest desert and dust source, with 9 million km² of nearly uninhabited land, whose area compares with that of the rapidly diminishing tropical rainforest. The objective would be to import moisture and reduce dust erosion in North Africa, followed by CO₂ absorption as new plant life develops.

How can new life be started in the desert? Pumping solar-desalinated seawater into selected basins could locally raise the humidity, lower the temperature, reduce dust erosion and encourage the founding of new photosynthetic colonies. The project would be costly, but it has several social advantages: (1) it would be the first 'biosphere project', (2) it is one answer to the question "Who will pity Africa?"¹¹, and (3) it would require the agreement and participation of seven or eight Saharan countries; acceptance of the project would, however, be helped by independent scientists with a goal of benefiting all of humanity. Nevertheless, global benefits would be slow in coming; several decades would pass before there were significant changes in humidity and CO₂. Dust erosion might yield sooner because strong erosion areas could be fed water first, creating many small lakes. Dust is a desiccant, so soil and humidity are both carried away by dust erosion. Combined with the sandblasting effect of high winds, dust is hostile to all forms of life.

One practical problem is returning brine to the sea, which could upset local ecosystems if not properly dispersed, but which is necessary because of the huge amount of salt involved. At least 1 per cent of the Saharan surface will probably need to be covered with water for an appreciable global effect. Supposing a mean depth of 10 m, this corresponds to >10¹⁰ tonnes of salt. Similarly, interaction of Saharan albedo, CO₂ increase, atmospheric warming and rainfall must be investigated further. Some experts think that a Sahara darkened by trees would contribute even more to warming because of reduced albedo. But such a forest would consume CO₂ and present a cooler surface to the air, counteracting albedo lowering; no clear synthesis of these opposing factors has appeared. Realiza-

tion of the project could start in individual countries with financing by an international agency, but developments that might be integrated into the overall effort should be encouraged. At the same time, catastrophes must not be permitted, as could occur if raw sea water were pumped directly into the desert. The optimal degree of desalination will depend on numerous factors. Many plants in the Sahara are halophiles and a wide range of degrees of desalting can be considered. Ground water may be available in parts of the Sahel, allowing a limited but more rapid realization.

This is not a proposal for competing for existing research funding. It is a request that funding for development of a corrective process be created. The long-term consequences of pumping large amounts of water into North Africa are difficult to assess and must be studied thoroughly in creating a model. Environmental parameters need evaluation, particularly new surface covered by water. Barring unforeseen events, we may have something like 50 years to consider taking control of the evolution of the biosphere. This supposes careful preparation for trying to correct the accelerating damage already done. The Sahara could become the world's greatest species park.

RICHARD GRANTHAM

*Institut d'Evolution Moléculaire,
69622 Villeurbanne cedex, France*

1. Beardsley, T. *Nature* **323**, 286 (1986).
2. Opinion *Nature* **321**, 713 (1986).
3. Folland, C.K., Palmer, T.N. & Parker, D.E. *Nature* **320**, 602–607 (1986).
4. Jones, P.D., Wigley, T.M.L. & Wright, P.B. *Nature* **322**, 430–434 (1986).
5. Loye-Pilot, M.D., Martin, J.M. & Morelli, J. *Nature* **321**, 427–428 (1986).
6. Prospero, J.M. & Nees, R.T. *Nature* **320**, 735–738 (1986).
7. Oerlemans, J. *Nature* **320**, 607–609 (1986).
8. Rasmussen, R.A. & Khalil, M.A.K. *Science* **232**, 1623–1624 (1986).
9. Lewin, R. *Science* **234**, 149–150 (1986).
10. Maddox, J. *Nature* **324**, 105 (1986).
11. Opinion *Nature* **321**, 548 (1986).

Learning the lesson

SIR—A recent report from the National Institute of Economic and Social Research, reviewed in the *Daily Telegraph* of 1 December 1986, highlights the growth of Italian industry over the past few decades. By no stretch of the imagination is it possible to show that Italian research lies at the root of this achievement.

I trust, therefore, that it is not too harsh to ask those who voice serious concern for the decline of our scientific effort, supposedly the mainstay of our industrial future, to seek the roots of the problem. All who dispassionately do so will find that it is commercial and industrial wealth that underpins scientific research and not the converse.

In today's competitive and highly tech-

nological world, it will be those nations that have attended to the education and the training of their youth that will create the wealth that was ours during the one and only industrial revolution. The most important contribution that our scientific community can make to the nation's future is to become wholeheartedly committed, today, to supporting a broad technical education throughout our schools.

D. A. A. FAGANDINI

*6 Alleyn Park,
Dulwich,
London SE21 8AE, UK*

Inflation run riot

SIR—Perhaps Marcello G. Cacace (*Nature* **324**, 508; 1986), in reading Paola De Paoli's short survey of Italian scientific research policy (*Nature* **323**, 751; 1986), was so troubled by the contrast between the lavishness of the 'million million' lire research projects and the smallness of CNR salaries that he did not notice that all the million millions were mistakes.

In fact, in De Paoli's article, 1,000 million lire is transformed into 1 million million lire with almost miraculous, but exasperating, consistency.

When reading the article at the time, I thought it showed a rare ignorance, among *Nature*'s staff, of the value of the Italian lira.

I now realize that the value of the Italian lira is unknown to Italians themselves, and I dread the moment at which the threatened 'lira pesante' will become a reality.

MARTA MONTELEONI

*22 Siddons Buildings,
Tavistock Street,
London WC2E 7NT, UK*

Classification

SIR—L.M. van Valen (*Nature* **323**, 664; 1986) is wrong in suggesting that provision of a helpful taxonomic service is incompatible with doing interesting research. Studies by taxonomists or others on the evolution of molecular biology of their organisms need not be used to evoke the nuisance of frequent changes in nomenclature.

In order to minimize this nuisance, codes of nomenclature should discourage such changes. A commission for each group of organisms should issue a definitive list of names at, say, quinquennial intervals. Apart from newly described organisms, this list would stand until the next list was issued. Taxonomists could continue to publish as hitherto, but formal acceptance of nomenclature changes would occur at intervals of several years only.

J.A. BARNETT

*School of Biological Sciences,
University of East Anglia,
Norwich NR4 7TJ, UK*

Where *Zitterbewegungen* may lead

There is a valuable place for journals a little to the side of orthodoxy that allow people to explore unfashionable themes. Even classical explanations of quantum mechanics have great interest.

In the 1940s and for some years afterwards, *Il Nuovo Cimento* was among the hottest of physics journals, if only because page charges (and currency exchange controls) prevented Europeans from using *Physical Review* except in the most exceptional circumstances. In those times, the Japanese had *International Journal of Theoretical Physics*, in which Tomanaga provided his distinctive counterpoint of the work of Schwinger, Feynman and Dyson on quantum electrodynamics.

Much of that changed with the founding of *Physical Review Letters* (by the American Physical Society and the late Samuel Goudsmit). *Il Nuovo Cimento*, now part of the European Physical Society's loose aggregation of journals, is now perhaps best known as one of the places in which Enrico Fermi used to publish. But it has also become a unique and valuable outlet for ideas on the edge of orthodox physics, and for that reason deserves more attention than it usually receives.

One recent issue, for example, contains no fewer than two articles taking as their starting-point the phenomenon of the "*Zitterbewegungen*", most tangibly that by means of which Pauli sought to provide a classical view of the origin of electron spin, but also a means by which people such as Schrödinger in the 1920s hoped to reconcile classical and quantum mechanics. *Zitterbewegungen* means 'shaking' or 'flickering' motion, and refers to the phenomenon by which particles subject to the uncertainty principle may be supposed to be constantly in motion on an unobservably microscopic scale. Several attempts to effect the reconciliation failed, yet by the early 1930s it was clear that the Schrödinger version of Dirac's relativistic equation for the motion of an electron implied that the instantaneous velocity of an electron could have only the velocity of light and its numerical opposite.

G. Cavallieri from the University of Brescia has for some time been carrying a torch for this approach to quantum mechanics; in 1985, he published such a derivation of Schrödinger's equation for a single particle. Now, with G. Spavieri of the Venezuelan University of the Andes, he has done the same for the many-body problem (*Il Nuovo Cimento* **95B**, 194; 1986) in a way that makes it possible to understand why previous attempts failed.

The starting point is the view that electron spin, and the magnetic moment that goes with it, is the electromagnetic con-

sequence of the random motions; unpolarized electrons move randomly in all directions, but the axes of the instantaneous circular motions of partly polarized electrons are restricted. Cavallieri likens *Zitterbewegungen* to the case of a rocket equipped with an infinitely large number of motors pointing in different directions which, by being fired at random, change the instantaneous direction of the velocity of the vehicle. It is essential to his case that this strictly inertial process is distinct from that, say, in Brownian motion, where changes of direction are brought about by interaction with the environment (an unsuccessful way of seeking a classical explanation for quantum mechanics).

The manipulation of these assumptions is uncomplicated. The trick is to calculate the probability density of an electron diffusing by these principles in some external field of force. The eventual result is Schrödinger's equation, but with the reservation that there is no way, at least for the time being, of calculating the numerical constants that arise, in particular quantities such as $\hbar/2m$ where $\hbar$ is Planck's constant and m the mass of the electron. The latest development, with Spavieri, is the extension of this calculation to a system of several identifiable particles.

Naturally, none of this amounts to a classical derivation of Schrödinger's equation. Even Cavallieri would not claim that, for the starting assumption of *zitterbewegungen* has no place in classical mechanics. In a sense, what this calculation has done is to lend a degree of tangibility to Pauli's old notion that would not otherwise have been just. Devotees of *Il Nuovo Cimento* will recognize this to be their journal at its most sober.

B. L. Cragin from the Center for Space Sciences at the University of Texas at Dallas goes to the other extreme (*ibid.* **95B**, 109; 1986). He starts with the observation by Dirac (*Nature* **139**, 323; 1937) that there are more "fundamental" constants than required to specify physical dimensions, with the consequence that it is possible to construct a set of dimensionless numbers from their ratios, but whose physical significance is unknown. The best known ratio is the fine-structure constant α (equal to $2\pi e^2/\hbar c$, where e is electron charge, $\hbar$ Planck's constant and c the velocity of light) which is, numerically, approximately $1/137$.

Cragin sets out to calculate what these numbers are from the assumption that the

microscopic *Zitterbewegungen* are also related to what are called the fluctuations of the vacuum in quantum electrodynamics. To put it crudely, the starting-point is the expectation that empty space is a kind of plasma consisting of the pairs of electrons and positrons that will be created whenever and wherever the fluctuations of the energy density of the field are sufficient to cause them to appear. For consistency, the microscopic motions with velocities $\pm c$ of each particle in the vacuum must both drive the plasma and be driven by it. There follows a classical calculation of the conditions that must be satisfied.

Sceptical readers will by this stage be asking "to what end?". Patience. The nub of the result is an equation roughly relating α^{-1} with $-(2/\pi) \log(2\alpha^2 \alpha_g^2)$ where α_g is the other dimensionless ratio $2\pi Gm/\hbar c$, numerically equal to 1.751×10^{-2} . And the outcome is a value for the inverse of the fine structure constant that is correct to five significant figures.

Perhaps one should not be too much awed by the accuracy with which the fine-structure constant emerges from these calculations, in the course of which Glazer has explicitly decided to neglect several factors such as the $4/3$ that normally precedes the usual expression for the volume of a sphere. He says himself that the agreement may be fortuitous. The crux of his argument is merely to be able to show some kind of a connection between the fluttering motion of an electron (and its partner positron) and the fluctuations of the vacuum even when there is no added energy (the zero-point fluctuations as they are called). Glazer is also anxious to make the point that his calculation of two non-dimensional ratios, one of which involves electric charge and the other the gravitational constant, is a sign that the same phenomena may link gravity with electromagnetism, a now-fashionable aim.

There are two views of these kinds of arguments, one of which is that, being founded on a bastard mixture of classical and quantum physics, they are merely means of exploring for consistency the dimensionality of the quantities involved. It is better than numerology, but, the purists would say, not much. But that may be too narrow. The arguments are entertaining, and they may have what is called heuristic value as well. Anyway, Cragin promises more, no doubt also in *Il Nuovo Cimento*.

John Maddox

Proofs

Are mathematicians logical?

Ian Stewart

IN A recent News and Views article (*Nature* 324, 406–407; 1986) I argued the case for rigorous proof in mathematics. But it would give a very one-sided picture of mathematics to lay emphasis on a technical necessity — proof — while ignoring the essence of the subject — ideas. So how do mathematicians really think?

The mathematical logician Alan Turing owned a bicycle, on which he cycled to work. At intervals the chain would fall off. A methodical man, Turing kept a bottle of turpentine and a rag in the office to clean his hands on arrival. After some time it dawned on him that the chain fell off at very regular intervals. On counting the revolutions of the front wheel he found exact regularity: after a certain number of turns, the chain fell off. He took to counting the revolutions so as to be able to execute a manoeuvre that kept the chain on. Tiring of this, he fixed a counter to the wheel. Later he analysed the mathematical relation between the number of spokes in the front wheel, the number of links in the chain and the number of cogs on the pedal; he discovered that the mishap occurred for a unique configuration of wheel, chain and pedals. On examining the machine he found that it happened when a particular damaged link came into contact with a particular bent spoke. The spoke was duly straightened and the turpentine and rag removed from the office. This tale illustrates both the power and the perils of logical reasoning. A cycle mechanic would have solved the problem in five minutes.

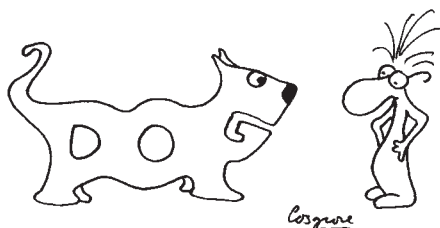
Textbooks on logic define a proof as follows. Assume a system of axioms and some logical rules of deduction. A proof is then a sequence of statements, each being either an axiom or a logical consequence of preceding statements, the final statement being the conclusion of the proof. When building up a theory the process is broken into stages called theorems, which may then be used in the proof of later theorems. This makes no essential difference provided we agree that the proof of any theorem tacitly includes those of all previous theorems used in it.

This formal view of rigorous logical reasoning is modelled on Euclid, and it was pushed to the fore by the nineteenth-century foundational crisis in mathematics. In practice nobody ever writes proofs this way, because life is too short. The description represents a logical ideal, and a proof is convincing if it is clear that 'in principle' a formal presentation of this kind would be a routine matter.

But this emphasis on formal structure

sets a snare: the assumption that a logical presentation of a mathematical idea is not just necessary, but sufficient. Mathematicians themselves have often encouraged this idea.

Perhaps the greatest mathematician of all time was Carl Friedrich Gauss. Another great mathematician, Niels Henrik Abel, says of Gauss: "He is like the fox, who erases his tracks in the sand with his tail". Carl Gustav Jacobi says that Gauss's proofs are "stark and frozen, so that one must first thaw them out". Gauss liked to polish the presentation of his ideas, rearranging them, erecting an elegant and logical edifice which often concealed the more intuitive ideas on which



they were based. He defended this practice: "When a fine building is finished, the scaffolding should no longer be visible". This attitude may encourage those who wish to admire the architecture, but it offers little help to those who wish to become builders themselves.

Another of the all-time greats was Archimedes. One of his most celebrated results is the formula for the volume of a sphere. He gave a rigorous proof by the classic Greek method of 'exhaustion', a technical device intended to get round difficulties involving irrational numbers. The big snag with exhaustion is that in order to use it, you must first have a pretty shrewd idea of the answer that you expect to get. Archimedes' proof — complicated but perfectly logical — is thus apparently plucked from thin air.

Archimedes did leave some hints about the plans of the building, if not the scaffolding itself. In one of his works he says: "Certain theorems first became clear to me by means of a mechanical method. Then, however, they had to be proved geometrically since the method provided no real proof. It is obviously easier to find a proof when we have already learned something about the question".

In 1906 came J.L. Heiberg's sensational discovery, in Constantinople, of a manuscript of Archimedes' lost work *The Method*. In it, Archimedes reveals the scaffolding. He guesses the formula for the volume by imagining a sphere to be

suspended from a balance, together with a cone and a cylinder. He cuts them all into infinitesimally thin slices, like a side of bacon, and rearranges the slices so that they continue to balance, but the masses of the cone and sphere become concentrated at one point. Applying the law of the lever and known formulas for the volumes of cone and cylinder, he derives the answer for the sphere.

Even in the time of Archimedes, then, mathematical research is a two-stage process. In the first stage, non-rigorous methods — indeed, downright ridiculous ones — are used to gain enough leverage to guess what the answer is. Then the ideas are reworked within a more rigorously structured framework of technique, to provide an acceptable proof that the answer is correct.

It is the same today. Intuition, analogies and a general sense of mathematical 'smell' combine to give an irresistible feeling that some particular piece of mathematics 'ought' to be true. It is then possible to begin the search for a technical proof. This second step is equally necessary: many results that ought to be true turn out not to be when examined in the cruel light of logic. But logic cannot operate in a vacuum: the ideas come first.

The mental processes of mathematicians have not been widely studied. The classic text, Jacques Hadamard's *The Psychology of Invention in the Mathematical Field*, dates from 1945. (It is available in a Dover reprint of 1954.) Hadamard notes that most mathematicians tend to work with vague pictorial images. In contrast he cites the case of the biologist Max Müller. When asked to visualize a dog, Müller declared his mental picture was

DOG

But Müller was an exception, even among biologists. And some mathematicians — traditionally the figure is held to be 10 per cent — also think purely in terms of the symbols on the page. Conceivably their thought-processes really are formal and logical, although I suspect they just have a different kind of intuition.

The most famous and informative inside account is that of Henri Poincaré, given in a lecture to the Société de Psychologie in Paris at the turn of the century. Poincaré has been described as "the last universalist", that is, the last great mathematician able to range freely over the whole of mathematics. (The subject has grown too big, rather than its practitioners too enfeebled, since.) He also took great interest in what a professional mathematician would consider to be more peripheral matters: popular exposition, philosophy, methodology. (How many of today's mathematicians would address a psychology society?) Of the logical approach, Poincaré says: "How is error possible in mathematics? A sane mind should not be guilty of a logical fallacy, yet there are

very fine minds incapable of following mathematical demonstrations. Need we add that mathematicians themselves are not infallible?

"The answer seems to me evident. Imagine a long series of syllogisms, and that the conclusions of each serve as premises for the following. Between the moment in which we first meet a proposition as a conclusion of one syllogism, and that in which we re-encounter it as the premise of another, occasionally some time will elapse... so it may happen that we have forgotten its meaning. We replace it by a slightly different proposition or attribute to it a slightly different meaning, and thus it is that we are exposed to error".

This is the danger of the logical step-by-step approach: we lose sight of the total picture. The point is made more succinctly in a passage from the writer Stefan Thémerson: "Only an elephant or a whale gives birth to a creature which weighs over 100 kg. The president's weight is 120 kg. Therefore the president's mother was either an elephant or a whale".

Poincaré also describes his experiences with a particular research problem. "For fifteen days I strove to prove that there could not be any functions like those I have since called Fuchsian functions. I was then very ignorant. Every day I seated myself at my work table, stayed an hour or two, tried a great number of combinations, and reached no results.

"One evening, contrary to my custom, I drank black coffee and could not sleep. Ideas rose in crowds: I felt them collide until pairs interlocked, so to speak, making a stable combination. By the next morning I had established the existence of a class of Fuchsian functions: I had only to write out the results, which took a few hours".

Poincaré concludes that mathematical creation involves four stages, which we might call preparation, incubation, illumination and verification. Preparation is a period of conscious work, often apparently fruitless. Incubation requires a period free of conscious thought, during which time the subconscious 'plays' with the problem. This leads to illumination: an idea strikes 'out of the blue' accompanied by a feeling of certainty. Finally, there is a period of conscious verification that the idea is in fact correct. (The process may, of course, get stuck or go wrong at any stage.)

Other mathematicians have had similar experiences. J.E. Littlewood relates that he had struggled for two months to prove a result that he was pretty sure was true. Half way up a Swiss mountain, a very odd device emerged — so odd that though it worked, he could not grasp the proof as a whole. He had a sense that his subconscious was saying "Are you *never* going to do it, confound you? Try this!"

Different mathematicians have different talents. William Thurston, at Prince-

ton, has been described as a "geometrical wizard". He waves his hands and wonderful things materialize from nowhere. Others, such as John Horton Conway at Cambridge, are especially good at unravelling combinatorial complexities. As a result, mathematicians spend a lot of time talking to each other, in the hope that one may see clearly through what to the other is impenetrable fog. I once witnessed a discussion between two colleagues about how to turn a sphere inside out. It lasted about an hour, was carried out over the telephone, and one of them was blind. They must have developed their own very personal set of images, but it was marvelous discipline. If you can explain topology over the telephone to a blind man you must understand it rather well.

The point that this series of anecdotes drives home is that the way mathematicians think about their problems is very different from the way that they present their formal results. It is the latter that has given mathematics a reputation for being

austere and dry, but it represents the surface, rather than the essence, of the subject. Although there may be advantages in appearing arcane and impenetrable, I rather think that the disadvantages carry more weight, the main one being that nobody outside the magic circle has the faintest idea what you are doing, or indeed that you are doing anything.

I am not suggesting that mathematicians start writing research papers including a potted history of all the blind alleys that they blundered into along the way. Journal space is finite. But it would be wonderful if some of the great mathematical minds of today would share with the rest of the world a little more of this inside story. If mathematicians can come to terms with the fact that their subject is the work of human beings, then this planet of human beings may become more favourably disposed towards mathematics. □

Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Low-temperature physics

Vortex production in helium

Peter McClintock

THE celebrated frictionless flow properties acquired by liquid helium at very low temperatures have long been known to vanish abruptly if the velocity of the liquid exceeds a certain critical value. The physical mechanism underlying this breakdown of the superfluidity has, however, remained an enigma, and it is pleasing, therefore, that a possible solution to this important and long-standing problem has now been proposed by K.W. Schwarz of IBM (*Physical Review Letters* **57**, 1448; 1986). The timing of his contribution is especially opportune in view of the current illuminating experimental work of E. Varoquaux (Université de Paris-Sud, Orsay) M.W. Meisel and O. Avenel (Centre d'Etudes Nucléaires de Saclay), some results of which are also published in a recent issue of *Physical Review Letters* (**57**, 2291; 1986). The latter authors are studying the onset of dissipation for an oscillating flow of superfluid helium through a very tiny orifice, using a technique which provides unprecedented sensitivity.

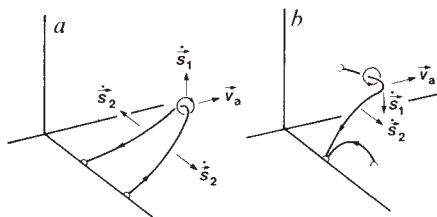
The basic experimental observation, which has been confirmed by many researchers, is that liquid helium at very low temperatures can pass at moderate speeds through tubes or orifices without any energy dissipation. The viscosity of the liquid is essentially zero, and so the flow can be maintained with no need of the pressure gradient that is required to keep a conventional liquid moving. But when the flow exceeds a characteristic critical

velocity, which usually seems to increase as the dimensions of the flow channel are reduced, there is a sudden onset of dissipation. It is generally accepted that this dissipation corresponds to the production within the liquid of quantized vortex lines, the quantum version of ordinary classical turbulence, by some unknown means. Any given element of vortex line can be envisaged as a narrow core region of atomic dimensions around which the superfluid flows at a tangential velocity that is inversely proportional to the distance from the centre of the core. Vortices within the liquid do not have free ends (which would be unstable) but, depending on circumstances, they can exist as isolated lengths stretched between the walls of the vessel; or as closed rings or other shapes within the body of the fluid; or as a dense tangle in which each line is eventually joined back on itself. Where pieces of vortex core approach, the flow field around each of them will influence the others and cause them to move.

In a remarkable experiment, Avenel and Varoquaux demonstrated convincingly (*Physical Review Letters* **55**, 2704; 1985) that the dissipative flow of superfluid helium through a very small rectangular orifice takes place in a sequence of discrete events. It is natural to suppose that these correspond to the appearance or movement of individual closed loops of vortex within the orifice. If so, one is bound to ask the fundamental question of whether they grow from elements of vor-

tex already trapped in the orifice (even in the absence of flow) or whether, on the contrary, each is created *ab initio* in a spontaneous transition within a vortex-free volume of the liquid — a very different kind of process.

Schwarz in his new work points out that a spontaneous transition is inherently improbable given the size of the orifice which, although microscopic ($5\text{ }\mu\text{m} \times 0.3\text{ }\mu\text{m}$), is nonetheless enormous in terms of atomic dimensions. Such a transition would clearly require simultaneous, coherent, changes in the motions of a huge number of atoms. It is more probable, he suggests, that the dissipation arises from the motion or growth of vortices already present in the liquid, as it is well established that even quiescent, non-flowing, superfluid helium contains a few such vortex lines. Both the expansion of vortex lines and their motion across the flow



Behaviour of pinned vortex lines in flowing superfluid helium-4, as deduced on the basis of a digital simulation (from Schwarz, K.W. *Physical Review Letters* **57**, 1448; 1986). *a*, Dissipation; *b*, vortex regeneration. In each case, v_a represents the bulk-flow velocity of the superfluid, and $\dot{s}_1$ and $\dot{s}_2$ represent the velocities with which local elements of vortex line will move. The ends of the vortex lines spend most of the time pinned to small protuberances on the walls of the flow channel as shown but, for sufficiently large values of v_a , they can escape, sliding along the walls until they get caught again by new pinning centres.

direction are dissipative processes. Furthermore, the effect of the flow on a loop of vortex pinned at its ends on small wall protuberances will be to make it expand and move as shown in *a* in the figure. Eventually, for a sufficiently fast flow velocity, the distorted loop 'de-pins' so that its ends can slide over the walls to 're-pin' on new sites, and the process repeats. The loop can thus travel dissipatively right across the channel. When it finally gets driven into the corners on the other side then, according to Schwarz's computer simulations, for sufficiently large flow velocities it re-orientates and re-connects (*b*) to form one or more small pinned loops that are suitably orientated for the initiation of new dissipative processes.

Schwarz suggests that the Avenel-Varoquaux observations can be accounted for if, for flow velocities just above those needed for de-pinning, vortices move across the orifice and then become pinned in the opposite corners in each half-cycle of the oscillating flow, so that they move across and back again once in each full

cycle. It is an ingenious idea, but further work is needed to establish whether it describes what actually happens. Although the mechanism provides a plausible explanation of many features of the experimental results, there are other features that are less easily encompassed by the model, the most notable of which is the observation by Varoquaux, Meisel and Avenel that the experimental value of the critical velocity increases linearly with decreasing temperature. There seems

to be no obvious reason why such behaviour should be expected on the basis of Schwarz's model. The experimental situation is not clear-cut, as the reproducibility of the data still has to be tested using different orifices, and the possibility of a microscopic leak in parallel with the orifice yet to be investigated. Further developments are awaited with interest. □

Peter McClintock is in the Department of Physics, University of Lancaster, Lancaster LA1 4YB, UK.

Microscopy

New ways to see atoms

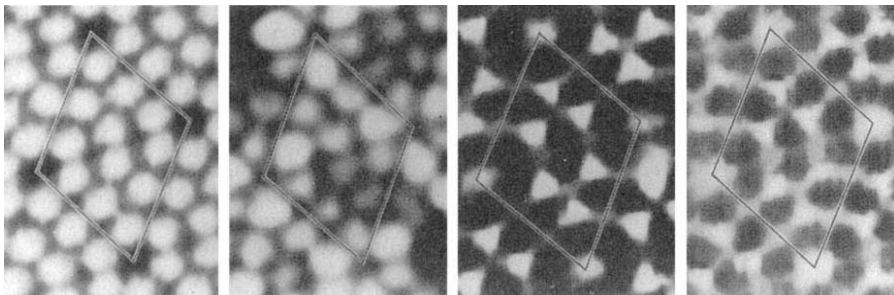
John Pethica

THE scanning tunnelling microscope developed by Binnig and Rohrer, who won the 1986 Nobel prize for Physics for this work (see *Nature* **323**, 663; 1986), is now well known for its ability to map the positions of atoms on surfaces¹. On semiconductor surfaces, simultaneous spectroscopy of the tunnelling current allows the real-space location of particular electronic states or bonds to be shown in a dramatic manner² (see figure). Two important extensions of scanning tunnelling microscopy have recently been established. First, the direct writing of nearly single-atom structure on semiconductors, which is described by J. Golovchenko and colleagues on page 419 of this issue³; and second, the atomic force microscope and imaging of the atomic structure of insulators.

The size of semiconductor devices is constantly being reduced in search of greater density of packing and of new physical phenomena to use in their operation. The limiting scale imposed by atomic structure now seems to have been reached for a memory device by Golovchenko and co-workers⁴, who show that a scanning tunnelling microscope can be used to 'write' a feature $8\text{ }\text{\AA}$ in diameter and about $1\text{ }\text{\AA}$ high onto an otherwise atomically smooth germanium surface. The presence of the feature can be subsequently 'read' using the microscope in its normal imaging mode. The mechanism

of writing involves raising the tip voltage of the microscope to about 4 V at the desired location. A distinct shift in apparent width of the tunnel gap indicates the point at which material is transferred from the tip to the germanium surface; the width of the gap remains stable on the surface as the voltage is then re-lowered. Structure within the written, disk-like feature has not yet been resolved, but its size and shape suggest that at most one or two atoms have been added to the surface and the immediate sub-surface region of the germanium. Curiously, it was not possible to make such a surface modification on silicon; evidently the stability and energetics of the surface atomic structure are critical. The mechanism of writing is not yet understood, but probably involves some form of field evaporation. It seems to be necessary to 'recharge' the tip by bringing it into contact with (a distant region of) the germanium surface before the writing process. Although many aspects of the process are not understood, it is likely that the smallest possible randomly addressable 'bit' of information has now been written.

The ability to vary the tunnel gap width in scanning tunnelling microscopy allows the determination of the electron decay length into vacuum, and hence of the tunnel barrier height. An early problem in the technique was that improbably low



Conventional scanning tunnelling microscope image (left) and the distribution of electrons in the various electronic states of the Si(III)-(7 $\times$ 7) unit cell (three right-hand frames) measured simultaneously. This technique allows the electronic structure of the surface to be mapped out in real space; in other words, the chemical bonds may be seen. From ref. 2, courtesy of Robert Hamers.

barrier heights (much less than 1 eV) were sometimes measured. The explanation⁴ is that forces could be acting between the tip and the surface which cause local elastic deformation and hence the tunnel gap movements to differ from the applied movements of the piezoelectric drive controlling the gap. These forces also have an effect on scaling in the microscope images. The idea of tip-to-surface forces acting in the scanning tunnelling microscope led Binnig, Quate and Gerber⁵ to suggest and to demonstrate the atomic force microscope. The principle of operation is similar to the scanning tunnelling microscope, but instead of tracing out contours of constant tunnel current, the microscope tip follows contours of constant force between tip and surface. The atomic force microscope thus closely resembles the traditional stylus profilometer, but the forces involved are several orders of magnitude smaller, in the region of 10^{-9} N; thus the area of interaction between tip and surface is very small. This low level of force is sensed by the deflections of a very soft spring on which the tip is mounted. There is evidence that at these extremely low loads, no permanent (plastic) deformation occurs to the surface, even on the single-atom scale, and therefore with a small enough contact area, atomic resolution may be possible.

In their first results⁵, Binnig *et al.* demonstrate a lateral resolution of about 30 Å. Recently (Ch. Gerber, personal communication) they have imaged the atomic structure of graphite, which suggests that they have attained a lateral resolution of about 2–3 Å. Resolution normal to the surface is, as for the scanning tunnelling microscope, in the sub-angstrom range. These results are important because the atomic force microscope overcomes one of the main limitations of scanning tunnelling microscopy — that the surface to be studied must be conducting and completely free of insulating contamination. Provided a good understanding of the mechanics of the tip-surface interaction can be obtained, in particular the role of longer-range surface forces and the types of deformation present, then a whole new field of study is opened up. Ceramics, insulating and other surface films, and biological materials of unknown conductivity (the usual case on the single-molecule scale) are a few of the possibilities. □

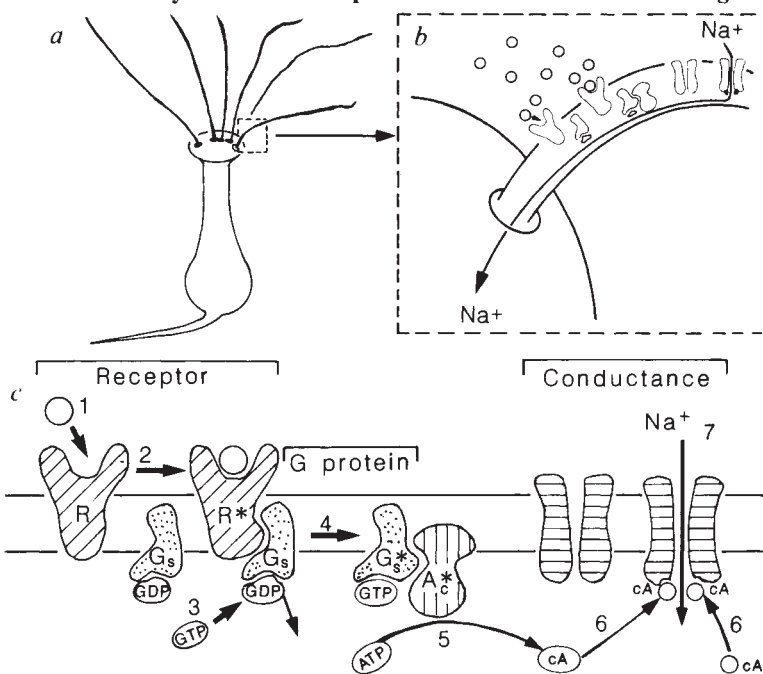
1. Binnig, G., Rohrer, H., Gerber, Ch. & Weibel, E. *Phys. Rev. Lett.* **50**, 120–124 (1983).
2. Hamers, R.J., Tromp, R.M. & Demuth, J.E. *Phys. Rev. Lett.* **56**, 1972–1976 (1986).
3. Becker, R.S., Golovchenko, J.A., & Schwarzenuber, B.S. *Nature* **325**, 419–421 (1987).
4. Coombs, J.H., & Pethica, J.B. *IBM J. Res. Dev.* **30**, 455–459 (1986).
5. Binnig, G., Quate, C.F. & Gerber, Ch. *Phys. Rev. Lett.* **56**, 930–933 (1986).

John Pethica is in the Department of Physics, University of Cambridge, Cavendish Laboratory, Cambridge CB3 0HE, UK.

Another cyclic-nucleotide-gated conductance

UNDERSTANDING olfactory transduction has been a sticky problem. As recently reviewed in News and Views¹, modern molecular and electrophysiological techniques are beginning to reveal details of the mechanism. On page 442 of this issue², Nakamura and Gold report that isolated membrane patches pulled from the cilia of olfactory receptor cells have a conductance that is gated open, apparently directly, by cyclic GMP or cyclic AMP. It seems probable that this conductance is the link between odorant-stimulated cyclase(s) in olfactory cells³ and the odorant-sensitive current (see figure), and indicates a close parallel between olfactory and visual transduction.

This discovery brings to four the number of distinct conductances that seem to be activated directly by cyclic nucleotides. Fesenko *et al.* first discovered⁴ this type of conductance in patches of plasma membrane from the outer segments of vertebrate rod photoreceptors, so providing the long-sought-for link between the cyclic-GMP cascade and the light-sensitive membrane current in rods, and completing the picture of the excitatory mechanism of phototransduction. Conductances gated by cyclic



a, The olfactory cilia bear receptor molecules and ion channels, shown enlarged in b. c, Biochemical reactions triggered by binding of olfactant molecule to the receptor (R), which converts it to active R^* (2). This binds to the G protein (G_s) causing phosphorylation of GDP to GTP (3) and converting G_s to G_s^* , which activates adenylate cyclase (A_c^*) (4) and promotes conversion of ATP to cyclic AMP (cA) (5). Cyclic AMP binds to and opens the conductance (6), allowing an inward current to flow, subsequently depolarizing the sensory cell (see b).

nucleotides have subsequently been found in cone photoreceptor outer segments⁵ and in the membranes of the ventral photoreceptors of the horseshoe crab *Limulus*⁶.

The four conductances are not, however, identical. The cone conductance is distinguished from that of the rods by its voltage dependence. The rod, cone and olfactory conductances are all very small, between 10 and 15 fS in the presence of normal divalent cations. The *Limulus* conductance, on the other hand, resembles a normal ion channel with a unitary conductance of 35–50 pS. The rod and cone conductances exhibit channel-like behaviour only in the absence of divalent cations. One prominent feature distinguishes the olfactory conductance from the others: it seems almost equally responsive to cyclic AMP and cyclic GMP, and to a much lesser extent to cyclic CMP², whereas the photoreceptor conductances all have strong selectivity for cyclic GMP and are relatively unresponsive to other nucleotides^{4,5}.

The progenitor of the family of cyclic-nucleotide-gated conductances may have already been present when simple ciliated receptor cells became specialized for either light or olfactory reception. The similarities that are emerging between photoreceptor, olfactory and hormone receptor mechanisms, all of which are mediated by GTP-binding proteins, are stimulating physiological, biochemical and genetic advances in the analysis of sensory transduction processes. We can expect rapid progress in unravelling the coupling between receptor and effector in these systems that are mediated by GTP-binding proteins and modulated by cyclic nucleotides. □

1. Shepherd, G.M., Getchell, T.V. & Mistretta, C.M. *Nature* **324**, 17–18 (1986).
2. Nakamura, T. & Gold, G.H. *Nature* **325**, 442–444 (1987).
3. Pace, U. *et al.* *Nature* **316**, 255–258 (1985).
4. Fesenko, E.E., Kolesnikov, S.S. & Lyubarsky, A.L. *Nature* **313**, 310–313 (1984).
5. Haynes, L.W. & Yau, K.-W. *Nature* **317**, 61–64 (1986).
6. Johnson, E.C. *et al.* *Nature* **324**, 468–470 (1986).

Transposable elements

Patterns of flower pigmentation

John R.S. Fincham

BARBARA McClintock, who discovered transposable genetic elements in maize, called them controlling elements and proposed that they were, or at least were derived from, part of the normal system of regulation of gene activity through development. Now that several such elements have been identified as DNA sequences, most of what McClintock deduced about their properties is being confirmed and extended at the molecular level, but her view of their normal controlling function is not generally accepted. The elements are widely regarded as examples of 'selfish DNA' — molecular parasites, albeit fairly harmless ones. Nevertheless, when integrated at certain strategic positions within the genome, they do have various intriguing effects on gene expression and are valuable as probes for the normal mechanisms of temporal and spatial gene regulation. The point is well made by work on transposable elements in the garden snapdragon (*Antirrhinum majus*) and especially in two recent papers from the John Innes Institute, Norwich (Coen, E.S. & Carpenter, R. *Trends Genet.* **2**, 292–296; 1986 and Coen, E.S., Carpenter, R. & Martin, C. *Cell* **47**, 285–296; 1986).

Investigations of genetic strains with spotty flower patterns have been carried out collaboratively between the John Innes group and H. Saedler's laboratory in the Max-Planck-Institut, Köln, resulting in the identification of three kinds of transposable element, Tam1, 2 and 3. These elements, like the known maize transposable elements, have inverted terminal repeats, resembling in this respect the *Drosophila* P (hybrid dysgenesis) factor and differing sharply from the other great class of transposable elements (exemplified by *Drosophila copia* and yeast Ty1 that have long direct terminal repeats and affinities with retrovirus genomes. Tam1 and Tam2, in fact, have some sequence similarity to the maize Spm (En) element, but Tam3, which is perhaps the most actively transposable of all, is not clearly related in sequence to anything else. It is with Tam3 that the latest publications are concerned.

Two *A. majus* genes are relevant to the story, each essential for the synthesis of cyanidin, the wild-type red-purple flower pigment. The normal allele of *nivea* (*Niv*⁺) encodes the enzyme chalcone synthase and the normal allele of *pallida* (*Pal*⁺) encodes dihydroflavonol reductase; these two enzymes respectively catalyse the first and final steps of the pigment pathway.

Plants recessive for null alleles of either of these two genes are devoid of red pigment. Plants with red-spotted flowers, not infrequently to be seen in gardens, most often result from an allele of *pal* called *pallida recurrens* (*pal*^{rec}) which is a null allele with an extremely high frequency of spontaneous mutation back to gene function. An earlier generation of John Innes



Array of *A. majus* flowers from very highly mutable plants heterozygous for *pal*^{rec} and a stable white allele showing overall dense spotting and larger sectors with various levels of pigmentation. (Photograph by the late L.S. Clarke, supplied by B. J. Harrison.)

geneticists (myself and B.J. Harrison, *Heredity* **22**, 211–224; 1967), following classical German work from the 1920s and earlier, had been particularly interested in the diversity of new *pal* alleles generated by mutation of *pal*^{rec}. Most were indistinguishable from standard wild-type *Pal*⁺ (uniform full-red), but a substantial number determined lower uniform levels of pigmentation ranging continuously from almost full-red down to barely tinged. Yet other *pal*^{rec}-derived alleles showed novel and reproducible patterns of pigmentation, with either weak or intense pigment confined to certain areas of the flower.

The John Innes–Köln team has shown that the bizarre properties of *pal*^{rec} can be attributable to a 3.5-kilobase Tam3 element inserted just upstream of the *Pal*⁺ transcription initiation site. The story of the capture of Tam3 in this location is an interesting example of the tactic of 'transposon tagging' in hunting for unknown DNA sequences (Martin, C., Carpenter,

R., Sommer, H., Saedler, H. & Coen, E.S. *EMBO J.* **4**, 1625–1630; 1985). Initially there was no molecular probe available either for Tam3 or for *Pal*⁺. But there was a complementary DNA clone from parsley corresponding to chalcone synthase messenger RNA, and this turned out to have sufficient similarity to *A. majus Niv*⁺. A search was made, in a particularly highly mutable *pal*^{rec} line, for plants exhibiting a mutable *niv* (*niv*^{rec}) phenotype, in the hope that such plants would have the element responsible for *pal*^{rec} transposed to the *niv* locus. Such did indeed turn out to be the case, and the parsley complementary DNA was used to identify a *niv*^{rec} genomic clone that included both *Niv*⁺ and a 3.5-kilobase Tam3 insertion a little distance upstream of the *Niv*⁺ transcription initiation site. From this clone, Martin *et al.* derived a Tam3-specific subclone which they used to identify and clone from *pal*^{rec} the Tam3 copy integrated just upstream of *Pal*⁺. The sites of integration of Tam3, just upstream of the two pigmentation genes, are indeed remarkably similar and suggest some degree of sequence specificity in the integration process. Be that as it may, the element in these two positions is ideally placed to interfere with gene expression in various interesting and instructive ways.

With a DNA probe for *Pal*⁺ now available, the way was clear to finding out what had happened to the *pal*^{rec}–Tam3 association in the different classes of *pal*^{rec} mutant derivatives. It was no surprise that the apparent *Pal*⁺ revertants appeared to have lost Tam3 from the *pal* locus — precisely so far as could be determined from restriction fragment sizes. What was more surprising was that several of the stable reduced-expression alleles that were investigated gave similar evidence of Tam3 excision. It is presumed that, in these cases, Tam3 has not been excised quite precisely but has left the *Pal*⁺ promoter region damaged in some way, for example by short deletions or insertions. Detailed sequence analysis of these 'scars' left by Tam3 should go far towards defining the relationship between promoter sequence and level of pigmentation, which, as Coen *et al.* were able to show, is, over a wide range, quite closely proportional to the level of *Pal*⁺ messenger RNA.

Even more interesting was the analysis of some of the mutants showing new stable patterns of pigment distribution. In three different mutants showing pigment in the corolla tube but not in the lobes, a not-quite-exact excision of Tam3 had left behind a modified sequence with the lucky easily recognized property of providing a cutting site for the restriction endonuclease *Bst*E11. It seems that this alteration in the promoter region has the effect of placing *Pal* under a form of developmental control to which it is not normally subject — at least not in the common garden

varieties of *A. majus*. The next, more difficult step, is to find out what it is that interacts with the altered sequence to modulate the gene expression.

Flower pigmentation obviously has a great advantage as a system for signalling gene expression; its self-indicating property makes unnecessary such complicated manoeuvres as fusing the gene under investigation to *Escherichia coli lacZ* and staining tissue preparations for β -galactosidase activity. The flower pigment is sufficiently cell-limited to indicate rather precisely the cells in which transcription of the pigmentation gene is switched on. We may expect soon to know

a good deal about the quantitative and developmental controls of both *pal* and *niv*, which are examples of unlinked genes whose activities are closely coordinated. Similar information may be expected on analogous systems in maize, where the genes *C₂* and *A₁*, homologous to *niv* and *pal*, respectively, exhibit a rich diversity of effects of inserted transposable elements, all documented in great detail by McClintock and now ripe for molecular analysis. □

John R.S. Fincham is Professor of Genetics at the University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

Host-parasite coevolution

The mite pockets of lizards

Michael J. Benton

THE study of parasite evolution, and in particular the coevolution of parasites and their hosts, is fraught with problems. The fossil record is patchy^{1,2} and it is hard to reconstruct the evolutionary story from living forms alone. An intriguing recently described example³, the case of chiggers and mite pockets in lizards, illustrates some of these problems and paradoxes.

Many lizards are infested by chiggers, the larvae of trombiculid mites, which feed on tissue fluid and cell debris. Surprisingly, lizards seem to go out of their way to attract the chiggers — they have special mite pockets that provide a protected, warm and humid site. In many cases, the skin of the lizard also has smaller scales than normal and a good blood supply in the pocket, which enables the parasites to feed more readily.

Arnold⁴ finds that chiggers are the most

common inhabitants of mite pockets, although ticks are occasionally found. The chigger is 0.25–0.5 mm long when it enters the pocket (see figure) and its size increases considerably after feeding on the host. It then drops to the ground where it metamorphoses into the adult form. Feeding involves the piercing of the skin and the injection of saliva through the stylostome, a tubular structure that develops in the skin of the host. The saliva causes cell breakdown and an increase in tissue fluid as a result of the inflammation reaction. The parasite may become fully engorged in a single day, or in a period of time of up to several weeks. Chiggers can cause severe damage to their hosts: lesions in the skin, allergic reactions to the saliva, secondary infection and transmission of disease organisms.

Mite pockets have been found in more

than 150 species and in 5 distinct families of lizard, which suggests that they have evolved independently many times. The pockets occur typically on the side of the neck, in the armpit region, high on the side of the chest, in the groin area and at the side of the base of the tail. Most species have pockets at only one location, but a few have two types of pockets. The pockets themselves range from slight folds in the skin to well-developed invaginations with a thick skin covering, often with reduced scales on the inside.

There are various models for the coevolution of parasites and their hosts. Brooks⁴ has divided the process in this context into two phenomena: co-accommodation, where parasite and host mutually adapt

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

British Museum (Natural History)

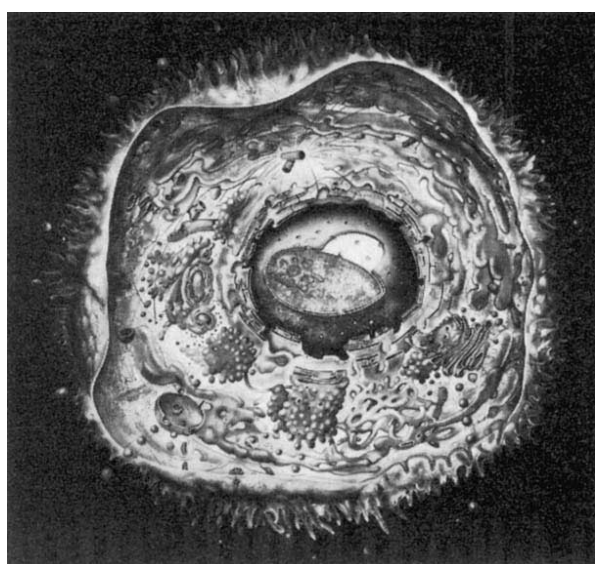
Chiggers in the axial mite pocket of the dwarf chameleon *Rhampholeon brevicandatus*. The hairy abdomens of the chiggers are about 0.5 mm in diameter (from ref. 3).

through time, and co-speciation, where a new parasite species arises either when speciation of the host occurs or when the parasite occupies a new host or takes on a new role in an old host. The relative importance of these two modes of host-parasite coevolution is not certain. In some groups, co-speciation seems to be the norm, particularly where parasites have specific host requirements and where co-accommodation is very narrow. If the host undergoes speciation, then the parasite probably speciates at the same time.

The chigger-lizard relationship seems to support another model of parasite-host coevolution in which co-speciation is of minor importance. Several ecologists⁵⁻⁸ argue that host-parasite relationships are largely the result of chance, analogous to the colonization of islands by free-living organisms. This approach suggests that the parasite load of any host is the result of a random 'first-come-first-served' process that depends on the rate of arrival of parasites and the rate at which they are rejected or die off.

Chiggers infest amphibians, birds and mammals as well as lizards; indeed, a single species of chigger can be found on members of three or four distinct host groups. *Trombicula* (*Eutrombicula*) *alfreddugesi*, for example, has been observed⁹ infesting 32 mammals, 52 birds, 39 reptiles and 3 amphibians. There is thus little evidence of close parasite-host specificity, nor of co-speciation. Chiggers seem to infest their

THE fascinating and complex world of the cell forms the basis of the newly opened Biochemistry Gallery at the Science Museum, South Kensington, London. The main attraction of the introductory exhibition *Cells, Molecules and Life* is a specially commissioned painting of the ultrastructural organization of the cell (right), the work of John Barber and Cynthia Clarke, which is linked to models illustrating various cellular processes. The gallery opened on 17 December 1986 and is



sponsored by the Biochemical Society, now celebrating its 75th anniversary; over the coming years the feature will expand to cover much of the biochemical basis of life. □

vertebrate hosts effectively at random.

This case raises another evolutionary problem: why should so many species of lizards apparently create structures to attract chiggers? Arnold³ finds that lizards with mite pockets are more likely to be infested with chiggers than closely related forms that lack pockets. This is unsurprising from the chigger's point of view, but why do lizards develop such structures?

Arnold's survey excludes the idea that the association of the chiggers with the mite pockets is fortuitous. The pockets are not a direct response to chigger infestation either, as they occur in newly hatched, uninfested individuals. But chiggers do apparently have several direct effects on lizards, including thickening of the skin around the stylostome and movement of lymphoid cells to the vicinity.

Arnold's best explanation for the mite pockets is that they are a form of damage limitation for the lizards. The pockets provide optimum conditions for chiggers and infestations are concentrated locally. The problem is contained, avoiding large-scale

disruption of skin and superficial organs. The skin in the pockets is resilient and recovers rapidly after the chiggers drop out.

There are parallels in other organisms — many plants have specialized galls or clumps of hairs where mites feed, and these structures apparently prevent the mites from damaging the photosynthetic surfaces of the leaves. Nevertheless, these kinds of structures seem to be an odd response to parasitism, and the risk arises that the parasites will take advantage of the superior feeding stations to such an extent that the host is overwhelmed. □

1. Morris, S. C. *Parasitology* **82**, 489 (1981).
2. Morris, S. C. & Crompton, D. W. T. *Biol. Rev.* **57**, 85 (1982).
3. Arnold, E. N. *Biol. J. Linn. Soc.* **29**, 1 (1986).
4. Brooks, D. R. *Syst. Zool.* **28**, 299 (1979).
5. Janzen, D. H. *Am. Nat.* **102**, 592 (1968).
6. Janzen, D. H. *Am. Nat.* **107**, 786 (1973).
7. Dritschilo, W., Cornell, H., Nafus, D. & O'Connor, B. *Science* **190**, 467 (1975).
8. Price, P. W. *Evolution* **31**, 405 (1977).
9. Wharton, G. W. & Fuller, H. S. *Mem. ent. Soc. Washington* **4**, 1 (1952).

Michael J. Benton is in the Department of Geology, The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland.

Geophysics

Mapping the mantle and core

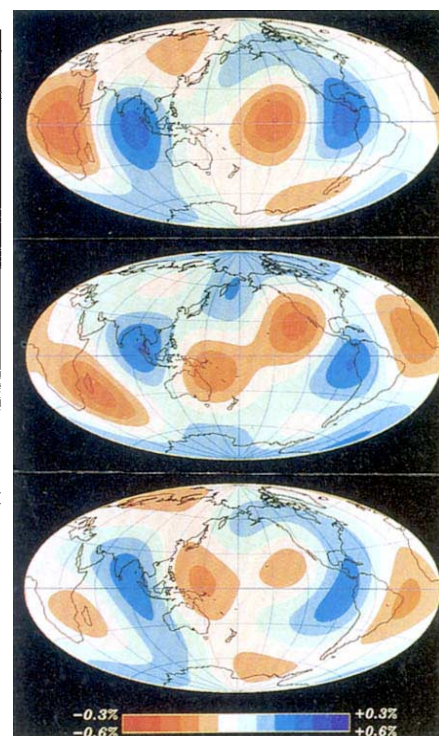
David Gubbins

It is common to talk of the effects of the plate tectonic revolution on geology; less so to speak of its effects on geophysics. It has left two obvious avenues of study open to the geophysicist interested in the study of the Earth's deep interior — either to view plate tectonics as a purely kinematic theory, describing only the convective motions of the Earth's mantle without addressing the question of why the plates move, and to try to supply a theory for the underlying dynamics of mantle convection; or to try to find new geophysical methods to map the deeper convective motions, again in a purely kinematic description. Several forthcoming papers, beginning with that by D. Giardini, Xiang-Dong Li and J. M. Woodhouse in this issue (*Nature* **325**, 405–411; 1987) suggest this latter approach is about to bear fruit.

The main tool for mapping the mantle is seismology. Seismic waves sample the elastic parameters, mainly the speed of compressional and shear waves, but also the density. However, a little knowledge of the properties of the material that constitutes the Earth's mantle and core can be used to convert seismic velocity into density and temperature provided the material is of uniform composition. Density, temperature and geoid (the equipotential surface of the Earth's gravitational field) anomalies can be combined with a model of mantle properties to determine the flow pattern, at least in principle (Hager, B. H. *et al.* *Nature* **313**, 541–545; 1985).

Adam Dziewonski (Harvard University) recently gave the Harold Jeffreys lecture to the Royal Astronomical Society. His theme was mapping the elastic parameters of the interior of the Earth — all the way from the central solid inner core to the surface — just as Jeffreys himself, with K. E. Bullen, had done almost 50 years ago. We can now map lateral changes in the elastic parameters, whereas Jeffreys was confined to a spherical Earth, layered like an onion. The long delay in progressing from the spherically symmetrical to the laterally varying model is a tribute to the careful work of Jeffreys and Bullen, and also an indication of just how difficult the extension has been.

Tomography, imaging of the interior of the Earth using seismic waves, should provide better information about shallow than deep regions, as earthquakes, the sources of the waves, are confined to shallow depths. Imaging the deep interior is like trying to look through frosted glass — the image is blurred by refraction in the glass. Seismic tomography attempts to remove the effects of the glass — the Earth's crust and lithosphere — using detailed models of the refractive index which has in turn been determined from tomography. The major pitfall is to map shallow structures into deep ones. Virtually all deep earthquakes occur in subducted plates, structures which exhibit the sharpest lateral temperature and consequent seismic velocity changes that exist anywhere in the



Three models of lower mantle structure of spherical harmonic degree 2 and 4. Top, P-velocity model V3 based on travel time residuals (± 0.3 per cent; Morelli and Dziewonski, 1986). Middle, P-velocity model based on splitting in free oscillation spectra (± 0.3 per cent; Giardini *et al.*, this issue). Bottom, S-velocity model based on SH waveforms (± 0.6 per cent; Woodhouse and Dziewonski, 1986).

Earth. This effect could dominate the true signal from the deep interior of the Earth, and the fear that this is the case has led many seismologists to be sceptical about the results from tomography.

In this issue, however, Giardini and colleagues present independent seismological evidence that is in broad agreement with the tomographic results. After a large earthquake the Earth rings like a bell. These elastic free oscillations have only been observed since 1960 but many of them have now been identified and their frequencies measured with great accuracy. Each frequency contains many modes, a degeneracy resulting from the spherical symmetry of the Earth analogous to the degeneracy of spectral lines from an atom. Departures from either spherical symmetry or isotropy will split the degeneracy and allow measurement of the fine structure of the spectrum. Observations of splitting are even more recent but can now be made routinely using the International Deployment of Accelerometers network of gravimeters run by the University of California at San Diego.

The new splitting measurements are directly sensitive to lateral heterogeneity and relatively insensitive to small-scale structure such as the subducted plates. It would be nice to report that Giardini *et al.* obtained the same solution for deep Earth structure from two quite distinct datasets,

but this is not what they have done: they merely show the two datasets are consistent with each other and, even so, there remain a few discrepancies. The authors also make the exciting suggestion that the solid inner core of the Earth is anisotropic, with axis of symmetry along the rotation axis of the Earth.

A. Morelli and A.M. Dziewonski (*Nature* in the press) have now mapped the surface of the fluid core using seismic waves that either travel through, or have been reflected from, the liquid core. This clever experiment evades the trap of mapping shallow structure onto the core surface. Reflected waves will arrive early if the core surface is shallow (if there is a bump on the core-mantle boundary) because it has less far to travel. A wave travelling through the same bump to the opposite side of the world will arrive late because waves travel slower in the core than the mantle. The effect of a bump on the core surface is therefore opposite for the two waves, but the effect of heterogeneity in the mantle should be the same. Good agreement is found between maps of the core-mantle boundary based on the two distinct datasets; it is unlikely to be caused by mantle heterogeneity and much more likely to result from true core topography.

Also, J. Bloxham and myself (*Nature* 317, 777-781; 1985) have mapped the magnetic field at the core surface for the past 265 years; we found permanent sites of stationary magnetic flux and other sites of persistent change in the magnetic field with time, as if the solid overlying mantle were controlling the field. This idea was suggested long ago by Raymond Hide but the theory could not be tested without information on the magnetic field and lower-mantle heterogeneity. We (Bloxham, J. & Gubbins, D. *Nature* in the press) have now found that these magnetic field sites are related to the mantle temperature measured by seismic velocity, and possibly also to the core-surface topography.

Other approaches are being used to address the same questions. For example, small changes in the rotation of the Earth, now measured very accurately by very long baseline interferometry, provide other stringent limits on the lateral heterogeneity of the Earth (Gwinn, C.R. *et al.* *J. geophys. Res.* 91, 4755-4765; 1986). Again, if the mantle controls the magnetic field then the same field patterns should be seen in the palaeomagnetic record for many millions of years. It is too early yet to predict another revolution in the Earth sciences, but the new results allow the hope that several different observations will be brought into line with a theory for whole-Earth convection. □

David Gubbins is in the Department of Earth Sciences, University of Cambridge, Bullard Laboratories, Cambridge CB2 1TT, UK.

Soft-hearted saurians in museum stores

The old adage, that many fossil discoveries are made in the museum basement, is confirmed by a recent study of ichthyosaurs, extinct marine reptiles like dolphins in form and habits, which highlights the often forgotten scientific importance of fossils in provincial museums¹. Leicestershire Museums have a collection of ichthyosaurs from Lower Liassic bituminous shales in the nineteenth-century quarries of the Barrow-upon-Soar district of Leicestershire. Their recent cleaning for display shows areas of exceptionally preserved soft parts such as skin, tendon and muscle or connective tissue. These areas, normally lost by decomposition before fossilization, are here preserved as very finely detailed mineral casts, masses of the bacteria and fungi which first decomposed the original tissue and then were killed and preserved by their own

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Leicestershire Museums

waste products, mainly the minerals calcium phosphate or apatite. The top figure shows a complete ichthyosaur skeleton, 2.54 metres long, with patches of soft tissue preserved around the neck, forelimbs and base of the tail.

These ichthyosaurs from Barrow-upon-Soar provide important new evidence for the anatomy of the soft parts of these reptiles, complementing the more famous specimens from the Upper Lias of Holzmaden in Germany². The soft tissues around the central bony skeleton of the forelimb show that the 'paddles' acted as hydrofoils, and have been used as evidence towards Jürgen Riess's radical reinterpretation of swimming in at least some ichthyosaurs: that they 'flew' underwater with their forelimbs like penguins, using the tail only for steering^{1,3}. It remains to be seen whether the Barrow-upon-Soar specimens can be used as evidence for some other model of swimming. But even if we continue to accept that the large tail fin of many ichthyosaurs was the main propulsive organ, this analysis reminds us that ichthyosaurs, like many other marine animals, had at least two 'gears': slow swimming by moving the forelimbs; and fast swimming with the tail fin. These fossils, clearly important, are now unobtainable. The source quarries have long been closed. Even if the quarries were open today, modern mechanical methods of excavation would cause the loss of many fossils that would have been recovered during the small-scale hand digging previously customary — the bottom figure shows the hand-worked Barrow-upon-Soar lime pits. The quarries at Holzmaden continue to produce their fine fossils only because the heritage law forbids mechanical excavation of the fossil-bearing strata, which are therefore dug by hand^{5,6}. So, in the absence of open quarries, the palaeontologist must excavate in the museum cellar.

Michael Taylor

1. Martin, J., Frey, E. & Riess, J. *Trans. Leicester Lit. Phil. Soc.* 80, 58-72 (1986).
2. McGowan, C. *Palaeontographica* A166, 93-135 (1979).
3. Riess, J. *Palaeontographica* A192, 93-155 (1986).
4. Wade, M. *Lethaia* 17, 99-113 (1984).
5. Keller, T. *Geol. Curator* 4, 193-198 (1985).
6. Wild, R. *Geol. Curator* 4, 275-280 (1986).

Michael Taylor was at the Area Museum Council for the South West, based at the City of Bristol Museum and Art Gallery, Bristol BS8 1RL, and has just moved to become Assistant Keeper (Earth Sciences), Leicestershire Museums, Leicester LE1 6TD, UK.

Social interaction

Survival in extreme isolation

Jared M. Diamond

WE think of ourselves as members of a social species for whom solitary confinement is the cruellest punishment, social isolation breeds psychopathology, and breakdown into small isolated groups leads to individuals becoming prone to malfunction. Interest in these issues is growing with concerns (for example, ref. 1) about whether any fragments of humanity left after a global nuclear war would survive the ensuing isolation. The best guides to these questions are several 'natural experiments' in prolonged extreme isolation, the longest of which ended 75 years ago.

On 29 August 1911, a starving, terrified, grief-stricken Indian man appeared at a corral in northern California, having just emerged from 41 years of hiding in a remote canyon. Ishi (as he became known) belonged to the Yahi tribe, who initially numbered several hundred until most were massacred by white Californian settlers between 1853 and 1870. In 1870 the 16 survivors of the final massacre went into concealment in the Mount Lassen wilderness and continued to live as hunter-gatherers. By 1894, their numbers had dwindled to five; by 1908 to four. In November 1908, surveyors stumbled upon the last Yahi's camp and took all their tools, clothes and winter food supply, with the result that three of the Yahi (Ishi's mother, his sister and an old man) died, leaving Ishi alone for three more years. Eventually he could stand it no longer and walked out to white civilization, expecting to be lynched there. In fact, he was employed by the University of California Museum at San Francisco, where he helped anthropologists to reconstruct his former lifestyle^{2,3}.

There have been a few documented cases of isolation similar to Ishi's, the most famous being the 18-year concealment of the HMS *Bounty* mutineers on Pitcairn Island (1790–1808)⁴. Less well known are the 18-year isolation of a lone Indian woman on San Nicolas Island (1835–1853)⁵ and the Japanese soldiers who remained concealed for decades in the jungle after the end of World War II on the Pacific islands of Guam, Morotai and Lubang^{6,7}. The marked differences among these groups or individuals in morale, health and death rate provide sobering lessons for prospective nuclear-war survivors. At one extreme, murderous dissension, suicide and alcoholism wracked the Pitcairn colony until 14 of the 15 adult men were dead. At the opposite extreme, the morale of the San Nicolas woman and especially of Lubang's last soldier remained excellent, even though they had to

endure complete isolation. If one compares the successful with the less-successful isolates, seven qualities notably absent in the *Bounty* settlers and present in the Yahi stand out:

1. Long familiarity with the local environment. Only the Yahi and the San



Ishi on 29 August 1911, the day he emerged from 41 years of hiding. (Lowie Museum of Anthropology, University of California, Berkeley.)

Nicolas woman had spent their entire lives in their outdoor prison and already knew how to find and extract its resources. In contrast, some of the Japanese soldiers starved.

2. Long harmonious experience as a group before isolation. The Yahi, unlike the other isolates, were a pre-adjusted group that had grown up together from infancy. The *Bounty* settlers were a mixture of three groups (from England, Tahiti and Tupuai) who met only months before sailing for Pitcairn.

3. Cultural homogeneity. The Japanese soldiers of Guam and Lubang, although they had not spent their lives together, had a common culture and language, whereas the Polynesians and English on Pitcairn had great cultural differences that were resolved by inter-racial murder.

4. No attempts by some group mem-

bers to parasitize others. Although all members of the Yahi and Japanese groups shared the work, some *Bounty* mutineers treated the Polynesians as slaves.

5. Non-competitive egos. Yahi were notably formal, polite and non-competitive, but sexual competition triggered a suicide and some murders on Pitcairn. Although the last surviving mutineer, Alexander Smith, is traditionally depicted as a saintly father-figure, he may actually have instigated the murders of other *Bounty* mutineers to get their wives⁴.

6. Willingness of individuals to assume new roles. The San Nicolas woman had to assume all the tasks traditionally associated with men; the Japanese soldiers assumed women's tasks such as cooking and sewing; and the Yahi men and women shared tasks. One seemingly trivial example that the last Lubang survivor cited to illustrate role flexibility was that when his group had dwindled to three, the procedure for haircuts was to rotate the roles of barber, client and sentry.

7. Shared belief in a righteous purpose. The Japanese soldiers on Lubang maintained their morale, even when only one survived, because they believed that the war was continuing and that they were fighting for Japan. They continued to light signal fires for Japanese ships that they hoped would be in the vicinity, even in the 1970s. Ishi's morale crumbled only after the deaths of all his companions had left him as the last survivor of a doomed race.

In the United States, groups called 'survivalists' have recently been practising how to live in rural wildernesses in the aftermath of a nuclear holocaust. How will they and the rest of us fare¹, compared with the Yahi? Post-holocaust fragments of humanity may well share the Yahi's advantage of cultural homogeneity, and the survivalists may be pre-formed groups. On the other hand, egos, attempts at parasitism, loss of sense of purpose and unwillingness or inability to broaden roles pose obvious risks. As for local knowledge, we may prove incapable of extracting metals, with the experts dead, the smelters destroyed and those ores accessible to simple technologies mined long ago. Who will be capable of rediscovering neolithic techniques that took *Homo sapiens* millennia to invent, and Ishi many decades to learn from his teachers? □

1. Clarke, R. *London Under Attack* (Blackwell, Oxford, 1986).

2. Kroeber, T. *Ishi in Two Worlds* (Univ. California Press, Berkeley, 1961).

3. Heizer, R.F. & Kroeber, T. *Ishi, the Last Yahi: a Documentary History* (Univ. California Press, Berkeley, 1979).

4. Kennedy, G. *Bligh* (Duckworth, London, 1978).

5. Hardacre, E. in *The California Indians* (eds Heizer, R. & Whipple, M.) 272 (Univ. California Press, Berkeley, 1971).

6. Onoda, H. *No Surrender: My Thirty-Year War* (Kodansha, Tokyo, 1974).

7. *Asahi Shimbun* correspondents *Twenty-eight Years in the Guam Jungle* (Japan Publications, Tokyo, 1972).

Jared M. Diamond is Professor of Physiology at University of California Medical School, Los Angeles, California 90024, USA.

Prediction of HIV vaccine

SIR—We read with interest the recent News and Views article¹ on the problems of producing an effective vaccine to HIV, the virus causing AIDS (acquired immune deficiency syndrome).

For some time we have been developing a novel computational approach for analysing protein sequences and predicting regions suitable for use as artificial vaccines. The program appears capable of recognizing not only polar sites, but also epitopic segments of predominantly hydrophobic character. The algorithms used will be published in full elsewhere, but, briefly, the method draws on information from secondary structure predictions, hydrophobicity analyses, scans for patterns of hydrophobic and hydrophilic residues consistent with loop and other structures, and the study of homologous sequences.

The procedure works well when tested on the available database of epitopic sites. Recently, the peptide of a steroid hormone receptor most strongly predicted to be a continuous epitope has been conjugated with a protein carrier and found to raise antibodies that recognize, with extremely high specificity, the nuclei of cells expressing these receptors. Based on this work, and as attachment of a carrier molecule is likely to be necessary for the immunogenicity of peptides and the conjugation procedure used may be important, we recommend that a short C-terminal extension ending in a cysteine residue be made, through which the peptide can be linked to the protein carrier.

We would be most interested to know if the regions of the HIV envelope proteins that we predict to be highly antigenic (Table 1) have been tried as synthetic vaccines, and if so, whether they are effective. This is particularly of interest as most of the strongly predicted epitopes in these proteins contain potential glycosylation

sites, especially in the putative extracellular domain of the transmembrane protein, whereas our predicted epitopes are a reduced set devoid of potential glycosylation sites, five of which correspond to predicted surface loops.

We are willing to use this program and other more elaborate methods^{2,3} to predict the antigenic regions of other proteins of possible interest in the production of an AIDS vaccine and to make the results freely available, if it is felt that they may be of value.

BARRY ROBSON

ROBERT V. FISHLEIGH

Department of Biochemistry
and Molecular Biology,

CHRISTOPHER A. MORRISON

Immunology Division, Department of Cell
and Structural Biology,
University of Manchester,
Oxford Road,
Manchester M13 9PT, UK

1. Newmark, P. *Nature* **324**, 304 (1986).
2. Robson, B. *et al. Int. J. Peptide Protein Res.* **25**, 1 (1985).
3. Robson, B. Platt, E. J. *molec. Biol.* **188**, 259 (1986).
4. Ratner, L. *et al. Nature* **313**, 277 (1985).
5. Meusing, M.A. *et al. Nature* **313**, 450 (1985).
6. Wain-Hobson, S. *et al. Cell* **40**, 9 (1985).

Protein structure and the specific heat of water

SIR—Studying the thermodynamics of the polymerization of collagen into fibrils, we have observed that the concentration of soluble collagen in equilibrium with fibrils is a minimum at 34.3 °C. Puzzled by the possible biological significance of this observation, we read with interest J. Paul's letter (*Nature* **323**, 300; 1986), in which he brought to our attention the fact that the specific heat capacity of water is minimal at 34–35 °C. To account for the apparent correlation between the temperature at which the specific heat capacity of water is minimal and the temperature at which higher organisms have evolved, he suggests that homiothermic animals having body temperatures around 35 °C have been selected because of their high efficiency at generating or dissipating heat in order to maintain their body temperature. Although the difference in specific heat capacity of water between 20°C (4.1819 J g⁻¹ °C⁻¹) and 35 °C (4.1782 J g⁻¹ °C⁻¹) is very small, it remains an attractive proposition that the evolution of warm-blooded animals and the chemical properties of water are related.

Is it feasible to assume that a change in the thermal properties of water is a consequence of a change in the structural organization of water molecules? If so, an alternative interpretation of J. Paul's hypothesis is that the structure rather than the thermal properties of water at 35 °C was the basis for evolutionary development of warm-blooded animals. It is interesting to postulate that the structure of

water at 35 °C is the most permissive to the evolution of the wide variety of biological structures that exist today.

Turning to our data on the solubility of collagen, 34.3 °C marks an abrupt transition in the thermodynamic properties of collagen in solution, and we would like to learn about other proteins and biological molecules whose structure/function appears to be responsive to the peculiar structure of water at 35 °C.

KARL KADLER

DARWIN J. PROCKOP

Jefferson Institute of Molecular Medicine,
Jefferson Medical College,
Thomas Jefferson University,
Philadelphia, Pennsylvania 19107, USA

EMBL database offer

SIR—The identification of newly sequenced proteins or parts thereof requires searches of protein sequence databases. May we inform your readers that we have undertaken the translation into protein sequences of the nucleotide sequences in the nucleotide sequence data collection of the European Molecular Biology Laboratory. Our database consists of 5,139 entries with a total of 952,078 amino acids. A tape with a flat file of our translated sequence data collection is available upon request.

The main features of our approach include the inspection of the features table, use of different genetic codes, splicing of exons as far as possible and control of proper reading frames by stop-codon test. In case of conflict, the entry has been omitted, as are sequences shorter than seven amino acids. Splicing was not undertaken if neighbouring exons appear in different entries.

RAINER STULICH

Zentralinstitut für Molekularbiologie,
Akademie der Wissenschaft der DDR,
Robert-Rössle-Strasse,
1115 Berlin-Buch, DDR

Mendel was no fraud

SIR—As a result of a paper by Fisher¹, and early discussion of the topic, the view has become widespread (for example, ref. 2) among those who have not bothered to read the biological literature, that Mendel falsified some of his data. Several analyses^{3–6} do not confirm this view, and results like Mendel's have even been found by others³. I wonder if some of the similar claims about other historic figures would also dissolve if suitably scrutinized.

LEIGH M. VAN VALEN

Biology Department (Whitman),
University of Chicago,
Chicago, Illinois, 60637, USA

Table 1 Antigenic sites predicted in the HIV *env* gene product

	Site	Sequence
Transmembrane protein	1	EESQNQQEKNEQE
Exterior protein	2	VPTDPNPQE
	3	PKVSFEPIPIHY
	4	KQSSGGDPE
	5	YAPPIISGQIR
	6	RPGGGDMRDN
	7	PTKAKRRVVQREKR

Sequences are given in IUPAC one-letter code. These predicted sites are identical in the three published sequences studied (4–6). It may be worth extending Site 3 to TQACP-KVSFEPIPIHYCAPA and closing the cysteines intramolecularly.

1. Fisher, R.A. *Ann. Sci.* **1**, 115–137 (1936).

2. Joyce, C.R.B. *Nature* **324**, 181 (1986).

3. Piegorsch, W.W. *Nature* **256**, 206 (1975).

4. Blixt, S. *Nature* **256**, 206 (1975).

5. Douglas, L. & Novitski, E. *Heredity* **38**, 253–257 (1977).

6. Monaghan, P. & Corcos, A.J. *Hered.* **76**, 307–309 (1985).

Testis size comparisons

SIR—J.M. Diamond's thoughts on the difference in testis size between individuals of European and Chinese descent¹ are accompanied by a pictorial representation in which the linear dimension is proportional to testis weight at autopsy. As proper scaling would have the linear dimension proportional to the cube root of weight, the size difference is vastly exaggerated.

Because drawings of human sex organs tend to catch the eye, casual perusers may have been left with the impression that pronounced testicular hypoplasia is the norm among males of Chinese heritage. Because sex-organ size is associated in the minds of some with courage, virility and the like, this careless illustration may initiate or promote a racial stereotype.

There has recently evolved into prominence a species of critic (for example, ref. 2) claiming that studies of intergroup differences are often hopelessly flawed. If their arguments have merit, it is important to avoid any hint of bias. I comment only because continuing interest in sex-organ size³ may portend future articles. If so, I hope the results are presented in tabular form. If pictorial presentation is unavoidable, the figures should be pre-pared with due regard for proper organ-scaling to eliminate any hint of an ethnic slur.

DAVID C. JOLLY

Box 931, Brookline,
Massachusetts 02146, USA

1. Diamond, J.M. *Nature* **320**, 488 (1986).
2. Gould, S.J. *The Mismeasure of Man* (Norton, New York, 1981).
3. Mittwoch, U. *Nature* **323**, 117 (1986).

More clues to gene rearrangements

SIR—The demonstration by Möröy *et al.*¹ of rearrangement of the *c-myc* gene in hepatocellular carcinomas of hepatitis virus-infected woodchucks deserves further consideration because the nucleotide sequences surrounding the rearrangement exhibit several features reminiscent of those involved in the rearrangements of immunoglobulin genes mediated by recombinase enzymes^{2,4}.

First, close to the breakpoint in the first intron of the woodchuck *c-myc* gene there is a heptamer–nonamer sequence motif which almost precisely matches that used by the recombinase⁵; the heptamer–nonamer is 19 bases 5' of the breakpoint rather than adjacent to it, as is usual, but these 19 bases may represent an N region⁶. Second, the rearrangement has occurred at a TG base pair in the *c-myc* intron, following an A residue on the juxtaposed DNA segment, consistent with the operation of an endonuclease, presumably associated with immunoglobulin recombinase activity, which cleaves at ^{TG}_{AC} sequences⁷. Thus, the breakpoint in this tumour shows characteristics similar to both physiologi-

cally rearranging immunoglobulin and T-cell receptor genes and chromosomal translocations in B and T cells involving the recombinase enzymes and *c-myc*^{2,4}.

Two explanations could reconcile these observations with the hepatocellular origins of the woodchuck tumours. It is possible that this tumour cell population consists of a B- or T-cell lymphoma, and that it is this which carries the rearranged *c-myc*. That would account for the observed oligoclonality of tumour W93, and the faint Southern blot rearrangement bands from tumours W74 and W64¹. Alternatively hepatocellular carcinoma cells might possess an enzymatic activity capable of recognizing sequences normally employed during immunoglobulin and T-cell receptor gene rearrangement. This activity might be endogenous to the hepatocyte that gives rise to the tumour, or it could be a result of hepatitis virus infection. We note that Hope *et al.*⁷ found the recombinase-like endonuclease not only in mouse fetal liver, a major site of B-cell lymphopoiesis, but also in adult liver.

FRANK G. HALUSKA
CARLO M. CROCE

The Wistar Institute,
36th & Spruce Streets,
Philadelphia, Pennsylvania 19104, USA

1. Möröy, T. *et al.* *Nature* **324**, 276–279 (1986).
2. Tsujimoto, Y. *et al.* *Science* **229**, 1390–1393 (1985).
3. Haluska, F.G. *et al.* *Nature* **324**, 158–161 (1986).
4. Finger, L.R. *et al.* *Science* **234**, 982–984 (1986).
5. Sakano, H. *et al.* *Nature* **286**, 676–683 (1980).
6. Tonegawa, S. *Nature* **302**, 575–581 (1983).
7. Hope, T.J. *et al.* *Science* **231**, 1141–1145 (1986).

MÖRÖY *ET AL.* REPLY — Haluska and Croce question whether the *c-myc* rearrangement could have taken place in infiltrating lymphocytes of neoplastic origin rather than in hepatocyte-derived tumour cells. Although Popper *et al.* have observed focal accumulation of haematopoietic cells in woodchuck liver tumours² and we observe lymphoid cells and megacaryocytes in these tumours, several arguments suggest to us that the *c-myc* recombination took place in liver tumour cells.

First, in the DNA of tumour W64, the rearranged *c-myc* bands are almost as intense as the normal bands (by exon 2 hybridization), while in W93 DNA they are even stronger than the normal band (see ref. 1, Figs 2b and 3b). It seems unlikely that the small proportion of lymphoid cells to liver cells in these tumours could account for the relative intensity of rearranged and normal *c-myc* bands on Southern blots of tumour DNA. Second, rearranged *c-myc* bands could not be detected in Southern blots of non-tumorous liver DNA, which contains the same population of lymphocytes as the tumorous livers. *In situ* hybridization is now in progress to identify the cell type producing the observed high levels of *c-myc* transcripts and therefore carrying a mutated *c-myc* allele. Finally, we have no evidence so far that the s/bg sequences¹ adjacent to *c-myc*

coding exons correspond to a part of the immunoglobulin or T-cell receptor loci. In contrast, these sequences hybridize with a 4.6 kilobase transcript, that is rather abundant in adult woodchuck liver, but absent from five other different woodchuck organs, including the spleen (unpublished results). Recombination of *c-myc* with cellular sequences specifically expressed in the liver argues for a localization of this event in liver cells.

We agree, however, with the hypothesis that the same enzymatic machinery might be involved in both hepatocellular carcinomas and lymphoid neoplasms. The analysis of *c-myc* rearrangements in other hepatocellular carcinomas may clarify the analogies noted by Haluska and Croce.

TARIK MÖRÖY
MICHELLE HADCHOUEL
PIERRE TIOLLAIS
MARIE-ANNICK BUENDIA

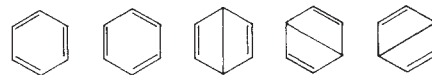
Institut Pasteur,
28, rue du Dr Roux,
75724 Paris Cedex 15, France

1. Möröy, T. *et al.* *Nature* **324**, 276–279 (1986).
2. Popper, H. *et al.* *Hepatology* **1**, 91–98 (1981).

Electronic structure of the benzene molecule

SIR—It is gratifying to me that a recent spin-coupled valence-bond treatment of the benzene molecule by Cooper, *et al.*¹ should have led the authors to conclude that the Kekulé description of the molecule, as expressed in the classical VB form, is much closer to reality than a description in terms of delocalized molecular orbitals.

The earliest reference given by Cooper *et al.* is 1961. In 1933, however, I published a simple graphical method of evaluating the matrix elements of the Rumer structures of molecules² and applied it to the five structures of benzene:



Solution of the secular equation³ gives the ratio of the occupation numbers of the Kekulé structures and the other three as 0.3898 : 0.0735, close to the values 0.4028 : 0.0648 obtained by Cooper *et al.*, who state that "They [their values] are astonishingly close to the values given long ago by Coulson⁴". Coulson, however, had in 1961 reproduced our 1933 calculation; the astonishingly close values had in fact been obtained in the earliest years of valence-bond quantum mechanics.

LINUS PAULING

Linus Pauling Institute
of Science and Medicine,
440 Page Mill Road,
Palo Alto, California 94306, USA

1. Cooper, D.L., Gerratt, J. & Raimondi, M. *Nature* **323**, 699 (1986).
2. Pauling, L. *J. chem. Phys.* **1**, 280 (1933).
3. Pauling, L. & Wheland, G.W. *J. chem. Phys.* **1**, 362 (1933).
4. Coulson, C.A. *Valence* 2nd edn, 249 (Clarendon, Oxford, 1961).

Behind the laboratory door

Hugh Rowell

Explorers of the Black Box: The Search for the Cellular Basis of Memory.

By Susan Allport. W. W. Norton: 1986. Pp. 271. \$17.95.

I STARTED this book filled with prejudice against it: I thought it probably combined muck-raking journalism with added publicity for an already over-exposed area. I apologize publicly for my suspicions. I will ask my graduate students and post-doctoral associates to read Ms Allport's work — they will find a great deal of value in it.

The author has written a very unusual book. I have an impression, based largely on reading book reviews, that historical accounts of research episodes and of the personalities and politics which shape them may not be uncommon in physics, but they are certainly rarities in biology. When they do occur, they are normally written by protagonists — one thinks of Watson's *Double Helix* — and are universally recognized as heavily biased renderings. Ms Allport, although a trained biologist, is not only an outsider to the area of research she describes, and thus with claims to objectivity, but she has also chosen to describe a still-developing episode, pursued by active scientists: she names all names, and she does not shrink from evaluations of personality, motive and character. Her chief characters include persons who devote exceptional efforts to obtaining publicity for their work, and can well be considered to have richly deserved the good and bad that it brings them — but I doubt whether many of them will be uniformly pleased by the image of themselves they see reflected here.

Ms Allport limits her study to a very small group of scientists and their work. They are the North American neurobiologists who in the past 20 years have busied themselves with the cellular basis of those functional changes in the nervous system of gastropod molluscs which are seen after exposing the animals to stimulus routines producing changes in the relation between stimulus and behavioural response: these workers characteristically describe the changes with terms originated by psychologists to describe outwardly similar effects in vertebrates, such as habituation, sensitization, associative conditioning, operant conditioning and the like. The author — like the scientists she writes about — ignores all related work on other groups of animals, vertebrate or invertebrate, or from other periods or other scientific communities.

With this tiny population she then attempts a very difficult feat, which has four components. First, she brings the

"general reader" enough cellular neurophysiology and learning theory to understand the work in question. Secondly, she gives a chronological account of the scientific progress of the various laboratories up to about the end of 1983. Thirdly, she portrays in detail the inter- and intra-laboratory rivalries and tensions, the disputes and the conflicting evidence, the cover-ups, the grantsmanship, the calculated absence of citations to the work of others, and the difficult questions of intellectual priority and of the influence over other scientists wielded by established workers with huge grants and with seats on most editorial and funding boards. Lastly, she attempts to portray the whole in a framework of the sociology of modern science, drawing heavily (and with due acknowledgement) on such authors as Robert Merton.

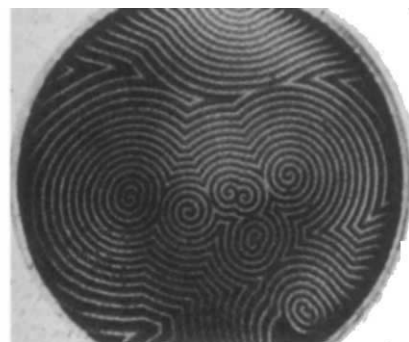
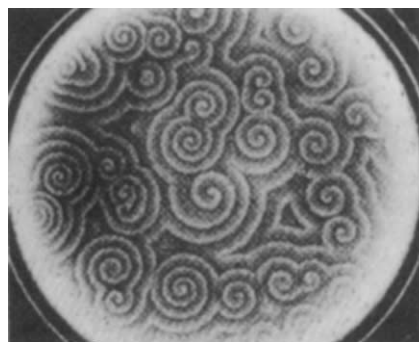
How well does she succeed? On balance, I would say that she does a very good job. Admittedly, the book has faults. The first third of it is unsatisfactory. There is no clear statement of aims, and the early, pedagogical parts are tedious. There the author often claims a scientific naivety — perhaps in an effort to strike a bond with the "general reader" — which the later chapters show her emphatically not to possess. There are also numerous anachronisms and historical errors — Lord Adrian is portrayed on p. 55 as using an "oscilloscope with a fluorescent screen" in the 1920s, and on p. 42 it is implied that chemical synaptic transmission was first conceived of after the introduction of the intracellular microelectrode. The distinction between these electrodes and extracellular microelectrodes as used in the cited work of Hubel and

Wiesel is not made. At one point, the crustaceans are said to be unsatisfactory for neurophysiology because they "die when they are opened up".

However once fairly into her account of the work and personnel in the laboratories of Kandel, Alkon, Davis, Lukowiak, Gelperin, Sahley, Pinsker, Byrne and others, things improve dramatically. Not only is the difficult subject matter uncomplicatedly expounded, her portraits and character sketches are always highly identifiable, sometimes hilariously or embarrassingly so, and her accounts of events fit well with what I myself know. Most importantly, perhaps, the author has grasped many fundamental points about research in neurophysiology in particular and biology in general that evade many researchers. She sees clearly the dangerous tendency of the medically trained worker to ignore the implications of biological diversity in both organisms and cellular mechanisms. She understands that at least some non-medically orientated workers on non-human nervous systems are interested primarily in learning about those systems, and not in extrapolating to the human state. She knows that the scientist must statistically expect his work to be erroneous or ill-conceived, and yet cannot admit this to the scientific or other public if his publications are to be taken seriously or if he is to be allowed to continue. She recognizes that power corrupts, not least in scientific practice. She can see that a field can approach Nobel Prize value in the public eye simply by dint of the amount of money and publicity expended on it, without having made more than normal progress.

The later part of the book is fluently written and fascinating reading. I could only wish that Ms Allport had emphasized more that the subfield she describes rides on the shoulders of the faster moving but less publicized research in molecular and membrane neurophysiology. □

Hugh Rowell is a Professor in the Zoologisches Institut, Rheinsprung 9, 4051-Basel, Switzerland.



Making waves — (left) wave fronts of cyclic AMP released by a monolayer of aggregating *Dictyostelium*; (right) a similar pattern, resulting from analogous kinetics, in the Belousov-Zhabotinsky chemical reagent. The illustration is taken from Arthur T. Winfree's *The Timing of Biological Clocks*, a new volume in the *Scientific American Library*. Publisher in Britain is W. H. Freeman, price is £15.95.

Palaeontology for the public

Tony Thulborn

The Dinosaur Heresies: New Theories Unlocking the Mystery of the Dinosaurs and Their Extinction. By Robert T. Bakker. *Morrow, New York: 1986. Pp.481. \$19.95.**

MORE than 12 years ago Robert Bakker provoked a heated controversy by insisting that dinosaurs were warm-blooded, intelligent and agile creatures bearing little resemblance to existing reptiles. Few scientists were entirely convinced by his arguments, though many conceded that there were probably some nuggets of truth among his more extravagant notions. Bakker's claims also roused some outspoken critics, who found his arguments as galling and insubstantial as a swarm of midges, and even today the subject is guaranteed to excite lively discussion — which usually ends in stalemate.

Now, in this fiercely partisan book, Bakker seeks to revive the great debate. Yet, surprisingly, he does not return to the fray with any new or more compelling arguments: instead he offers a catalogue of casual opinions and grand generalities, while his underlying arguments recede further into the background or are dismissed as briefly as possible. In essence the book is an inflated compendium of Bakker's ideas about dinosaurs, garnished with exuberant illustrations.

Despite the publisher's hyperbole, this book is not "a full-fledged, highly respectable work of scholarship": it is rhetorical rather than descriptive or analytical, and contains relatively little in the way of hard facts. It is also unashamedly one-sided in its presentation of ideas. Criticisms of Bakker's work (and there have been plenty of them) are simply shrugged aside or ignored entirely, whereas Bakker's own views, however outlandish, are presented as logically inescapable realities. So, for example, the brontosaurus must have given birth to live young because they were "too big to be laid in eggs" (p.357); the Triassic reptile *Lagosuchus* is "a good candidate for proto-pterodactyl" because it was lightly built and had its neck curved into an S-bend (p.293); and predatory dinosaurs such as *Ornitholestes* "succeeded in keeping the Mammalia small for over one hundred million years" (p.98). There are scores of such contentious propositions and speculations, all unblushingly presented as sound scientific opinion.

The book is written in an exaggerated

style, studded with superlatives to end all superlatives. Dinosaurs are not merely impressive or awesome, they are "super-huge" or "super-giant", and they constitute nothing less than a "megadynasty" that suffered "mass murder" at the end of the Cretaceous. Numerous line-drawings depict these "perfectly adapted masters of the terrestrial ecosystem" in suitably energetic postures: the three-ton stegosaurus *Diracodon* teeters precariously on one foot, and the bird-like theropod *Deinonychus* squabbles amid a flurry of feathers. The splendidly imaginative illustrations are Bakker's own work, and the book is worth purchasing for these alone.

As Bakker admits, many of the theories he presents are developed from ideas in the older literature and in the works of extant authors (who are not always identified by name). But my greatest grumble is with Bakker's grotesque caricature of palaeontology and its practitioners. Those

who don't happen to share Bakker's views about dinosaurs appear to be bumbling and unimaginative drones, so devoid of intellectual activity that they are barely distinguishable from the fossils they study. And, between the lines, a gullible reader might gather that sound palaeontological theory may be cobbled together by anyone with a smattering of knowledge and a vivid imagination. Bakker stresses that palaeontology is a hard science, but his style of argument would scarcely convince anyone of that fact.

Palaeontologists will probably react to this book with amusement, incredulity or anguish, depending on the extent of their professional involvement with dinosaurs. Non-specialists are likely to deem it a *tour de force*, though they should be warned not to swallow too many heresies without an occasional pinch of salt. □

Tony Thulborn is a Senior Lecturer in the Department of Zoology, University of Queensland, St Lucia, Brisbane, Australia 4067.

Storage facilities

Daniel L. Schacter

Working Memory. By Alan Baddeley. *Clarendon: 1986. Pp.289. £30, \$45.*

IN THE late 1960s, one of the proudest achievements in all of cognitive psychology was the emergence of a detailed theoretical account of an entity known as the short-term store — a kind of mental buffer system that could hold a limited number of items for brief intervals before they were transferred to long-term storage. Theories of the short-term store were often characterized by impressive mathematical precision, and they converged upon a common view that came to be known as the modal model. Some psychologists talked enthusiastically — and in retrospect, perhaps naively — of focusing theoretical attention on problems of long-term memory because the problem of the short-term store appeared to be solved. Then the roof fell in. Within the span of a few short years, experimental data appeared that were inconsistent with the modal model, compelling arguments were advanced that seriously undermined the very notion of a short-term store and the once orderly field of short-term memory was plunged into theoretical disarray.

All was not lost, however. In 1974, Alan Baddeley and Graham Hitch published an article that suggested a new agenda for research on short-term memory. Their agenda consisted of two key items: a theoretical framework that attempted to overcome problems associated with the modal model, and a general approach towards the study of short-term memory processes in a range of cognitive activities,

including comprehending, reading and reasoning. In *Working Memory*, Baddeley reviews and assesses a decade's worth of research that has been pursued in the context of the agenda that he and Hitch set out. Judging by the material in this highly readable and comprehensive monograph, that agenda has proved to be an exceptionally fruitful one.

One of the main differences between the modal model of the 1960s and Baddeley's working memory model is that the former depicted the short-term store as a unitary entity, whereas the latter fractionates working memory into several components. The critical components are the central executive, a sort of general-purpose processor which is involved in decision, selection and control functions, and two subordinate or slave subsystems — the articulatory loop, which is crucial for temporary storage of speech-based information, and the visuo-spatial scratchpad, which can hold and represent a limited amount of visual and spatial information. More generally, Baddeley advocates that this system, or something like it, operates as a kind of mental workspace where information can be temporarily held and used in the service of performance on a wide variety of cognitive tasks.

Most of the studies carried out by Baddeley and others within the framework of working memory use a similar experimental strategy. Subjects typically perform a *primary* task, which may involve reading, comprehension, problem-solving or reasoning. At the same time, they also perform a *secondary* task, such as rehearsing and then recalling a small set of digits, which is presumed to draw on the capacity of working memory. The major question is whether performing the secondary task disrupts performance of the primary task.

*To be published in Britain by Longman with the subtitle *A Revolutionary View of the Dinosaurs*. Provisional price is £14.95 and month of publication April.

If it does, the inference is that working memory is involved in performance of the disrupted primary task. Using this basic strategy, Baddeley and others have produced a wealth of research that has delineated the cognitive functions of the various components of working memory, clarified a number of theoretical issues and provided new perspectives on cognitive impairments in various populations, including learning-disabled children and patients with Alzheimer's disease.

The principal value of *Working Memory* is that it brings together in one place the best of this research, and it does so effectively. The author's portrayal of his own work and that of others is consistently incisive, critical and even-handed. If there is one disappointment, it is that Baddeley does not address a problem that seems so central to his subject that it cries out to be confronted head on: the place of consciousness in working memory. Baddeley's reluctance to delve into this issue may stem from the relative lack of research on the central executive component of working memory, which is where questions concerning consciousness arise most directly. In fact, the author acknowledges his own disappointment with respect to our ignorance of the central executive, and discusses some promising ideas that may help to reduce it.

In many areas of psychological inquiry, research fashions change with alarming swiftness; sustained programmatic investigations of key issues are the exception rather than the rule. *Working Memory* is a solid demonstration that experimental rigour and ingenuity, coupled with a broad conceptual approach and sheer persistence, can result in something that is as rare as it is invaluable: genuinely cumulative progress. Cognitive psychology owes Baddeley and his associates a debt of gratitude. □

Daniel L. Schacter is an Assistant Professor in the Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1A1.

New in paperback

- *Functions of the Brain* edited by Clive Coen. Publisher is Oxford University Press, price is £8.95, \$24.95. (Reviewed in *Nature* 320, 223; 1986.)
- *Radiant Science, Dark Politics: A Memoir of the Nuclear Age* by Martin D. Kamen. Publisher is University of California Press, price is \$8.95, £7.50. (*Nature* 318, 607; 1985.)

New in the United States

Two books published in Britain and reviewed recently in *Nature* are now available in the United States:

- *False Prophets: Fraud, Error and Misdemeanor in Science and Medicine* by Alexander Kohn. Publisher is Basil Blackwell, price is \$24.95. (Reviewed in *Nature* 324, 181; 1986.)
- *The Red and the Blue: Cambridge, Treason and Intelligence* by Andrew Sinclair. Publisher is Little Brown, price is \$17.95. (*Nature* 322, 217; 1986.)

London learning

W. F. Bynum

The University of London 1836-1986: An Illustrated History. By Negley Harte. *Athlone, London and Atlantic Highlands, New Jersey: 1986. Pp. 303. Hbk £11.95, \$39.95; pbk £4.95.*

HUNDREDTH birthdays often warrant a letter from the Queen. The sesquicentennial of the University of London gets instead some kind words from Princess Anne, its Chancellor since 1981, to preface Negley Harte's splendid illustrated history of the institution. The 366 pictures and their informative captions are part of his volume, not simply gratuitous adornment of it.

To many, the University of London is a nebulous bureaucratic entity which lives in 'Senate House', that monstrosity just north of the British Museum which would look more at home in Red Square, Moscow, than Bloomsbury. To others, the University is confused with University College, founded after all a decade before the University of London and originally bearing that name. Still others might be tempted to echo Professor Ralf Dahrendorf, former Director of the London School of Economics (yet another part of the federated University): "You mean the so-called University of London". Whatever the misconceptions one brings to Harte's volume, they will be dispelled. His considerable achievement is to have clarified a complex story, and to have turned what could have been a heavy and dull account into an entertaining and witty narrative.

The occasion which makes 1986 a sesquicentennial was a charter which provided *inter alia* an examining and degree-awarding structure for University College (as it then became) and King's College, London, an Anglican establishment founded as a counter to the essentially secular character of the original 'University of London' in Gower Street. Examining has always been a prime function of the University, and from the earliest days candidates could present themselves for examination even if they had not studied in London. By this procedure, the University exerted an enormous influence throughout Britain and her Empire, because it offered the opportunity for individuals — from Manchester or Cairo, Sydney or Bombay — to obtain a degree. Even today, the external degree system continues as a vital part of the University's activities, and it was through this that the curriculum was expanded beyond the traditional classics and mathematics available at Oxford and Cambridge. Science was an early beneficiary of this enlightened policy; so, at a slightly later date,

were women who, after 1880, were eligible for London degrees.

In addition to its examination service, the University has provided a structure for coordinating higher education in the Metropolis. This federal University has become one of the largest in the world, consisting of 24 'Schools' (for example University College, Imperial College and the London School of Hygiene and Tropical Medicine) and 13 'Institutes' (for example the Institute of Education and Institute of Historical Research). Relations between

Pilgrims from many paths we came
To where the roads of Empire meet,
Our lives to kindle at the flame
Of schools wherein a million feet
Have trod the years, or with a fame
That yet along the years shall beat.

O London maids and London men
Bring in the golden age again.

In no seclusion pastur'd round
As where the Cam and Isis flow,
Our cloister'd learning have we found,
Where loud the tides of traffic go.
Our nightingales have been the sound
Of London bells from Fleet to Bow.

O London maids and London men
Bring in the golden age again.

Life calls us, and we bid farewell
To this the latest of our springs,
But on our travels we will tell
How fellowship of gentle things
Is kept for ever where they dwell
Who know the song that England sings.

O London maids and London men
Bring in the golden age again.

In field or market place or mill,
Beneath a dear or alien Sun,
We'll build a generation still
Of faith and honour here begun,
That sires of the old English will
Shall know their own and cry: Well done.

O London maids and London men
Bring in the golden age again.

*Echoes of Empire — "Graduation Song"
of the University, first performed in 1926.
It was soon dropped.*

the various bodies within the University have not always been harmonious, since their several histories and functions are so diverse. Indeed, even the present loose federation may not be viable if the economic climate continues to deteriorate.

Certainly, with 'rationalization' being one of the key words of the 1980s, the University's future must be under a cloud. Those who believe in it can take comfort from the fact that its future has always been uncertain, and that it has managed to survive innumerable charters and commissions. In the meantime, anyone interested in higher education has a treat in store, for Harte's book makes sense out of Senate House. □

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BP, UK.

Light on lasers — and a lot of it

C.R. Pidgeon

Lasers. By Anthony E. Siegman. *University Science Books, 20 Edgehill Road, Mill Valley, California 94941, USA/Oxford University Press, UK: 1986. Pp.1,283. \$58, £60.*

In 1971, in *An Introduction to Lasers and Masers* (published by McGraw-Hill), A.E. Siegman described lasers and masers in the context of electronic amplifiers; the main differences of these devices from earlier ones such as transistors and tunnel diodes is that the amplification is through interaction between electric (and magnetic) fields and the internal resonances of atoms and molecules. The subject was described very successfully in terms of the semi-classical concepts of the electron oscillator model, providing a textbook suitable for first-year graduates either in physics or electrical engineering.

Since 1971 the subject has continued to proliferate and we now have the present, very impressive volume, which runs to 1,283 pages compared to the 520 pages of its predecessor. An important reason for alluding to the previous book is that although Siegman now goes into the subject at considerably greater depth, and over a much wider range, his approach remains the same: "to remove much of the quantum mystique from quantum electronics"; the classical electron oscillator results are converted to quantum mechanical results for real atomic transitions.

Siegman's prescription is, where necessary, to replace the total number density of the classical oscillator by the population difference density of the actual atomic transitions involved; the linewidth is interpreted as that associated with the atomic transition, and the radiative damping time with the atomic transition lifetime (or the inverse of the quantum mechanical transition rate). A normal rate equation analysis then follows. In the quantum case the population difference becomes a function of time, which means the rate equation limit fails for strong enough applied signals and leads to Rabi flopping behaviour. The nonlinear polarization equation has to be solved together with a time dependent population equation. The general subject of coherent transient phenomena is also dealt with (towards the end of the book), by allusion to the Bloch vector equations for magnetic dipole transitions.

As the author points out, there is a large amount of material here. But it is conveniently divided into sections that can be used at all levels from introductory graduate work to an advanced text for practising

laser engineers and research workers. As may have been anticipated by those attending laser Summer Schools over the years, there are excellent chapters devoted to such topics as optical resonators, Q-switching, mode locking and injection locking, which are a complete guide to the current state of research. Other areas are well referenced for further reading. Where necessary, laser systems are described in detail to illustrate physical principles, but no attempt is made to be encyclopaedic in this regard. Since the author comes from Stanford University one is tempted to ask, what about the free electron laser? It doesn't get a mention!

This is an excellent book, and a very

good buy at \$50 (if less so at £60). The only quibble I have is that, once one breaks the 1,000-page barrier, is it really worth all the effort to avoid what is the most elegant and logical description of some of the coherence phenomena — namely the density matrix model? After all, to quote the author's own comparison with the understanding of solid state devices as adequate in terms of a "classical" description of electrons and holes; this does not, for example, explain what tunnelling means in tunnel diodes. However, I do agree that we don't need to hear, *ad nauseam*, about perturbation theory. □

C.R. Pidgeon is a Professor in the Department of Physics, Heriot-Watt University, Edinburgh EH1 1HX, UK.

Talk of nucleotides

Richard Grantham

Gnomic: A Dictionary of Genetic Codes. By Edward N. Trifonov and Volker Brendel. *Balaban Publishers/International Science Services, POB 2039, Rehovot, Israel: 1986. Pp.272. Hbk \$80; pbk \$60.*

In *Gnomic*, Trifonov and Brendel seek to help us understand the "language" of nucleic acid sequences. Their tome of sequence "gnomology" (knowledge of oligonucleotide meaning) suggests several questions: Who could use it? Is it up to date? Is it, as intended, a complete pan-genetic dictionary?

The authors have assembled about 800 "words" of pithy meaning in genetic sequences; in their lexicon, each oligonucleotide-word is accompanied by note of its biological meaning or function, with appropriate references. Such an aid for bench scientists studying promoters, operators, protein-binding sites and so on could soon find service since the huge second-generation sequence banks are on the horizon. However, the book is biased and incomplete. The first main section, "Gnomic Dictionary", does not give authors' names or dates of publication of the papers concerned, and some journals are favoured over others: *Science* is cited only eight times, the last reference being from 1983 when that weekly was starting to publish more molecular biology; *Nature* is listed six times more often. Next comes the "Context Index", revealing in which published sequences each of the 1,024 possible pentanucleotides has been found. Many of these pentamers are absent in our biological system if we can believe the data here; AAGCG, AAGGC, ACGCC, ACGTT, AGACG, AGATT, AGCAC and over 50 others apparently do not exist, but this is more than likely a reflection of the prematurity of the text.

The third major section, "Keyword

Index", shows entries for intron, goat and beta-globin, but none for exon, sheep (or beef) and alpha-globin. *Gnomic* could not be up to date, because the number of sequences published nearly doubles every year and there is no end in sight. But the authors indicate a workbase of only six million letters, and their already obsolete sample thus averages probably about one nucleotide per extant species. The book is also incomplete through neglect of references; the 400 papers screened are simply too few! Indeed, Trifonov and Brendel are excessively auto-referential and ignore pertinent analyses by other groups in their discussions of the "Morphology of Gnomic" (by which they mean preferential use of certain words in composing a genetic text). Finally, the appendices are hampered by unclear explanations and outdated data. Appendix IV, "Codon Usage Tables", is just a jovial blank page with a reference.

This book requires extensive rethinking and updating before it can appeal to researchers and students. For example, the present double reference system is unnecessarily cumbersome; full citations with standard journal abbreviations could appear in the "Gnomic Dictionary". And, of course, the references could be put in alphabetical order by author name, becoming an author index (missing at present) and slightly humanizing the book's contents.

Trifonov and Brendel are right in encouraging us to do more comparative analysis and interpretations on the sequence information available, but bench sequencing will probably continue to dominate because of our preoccupation with biotechnology (*nostra culpa*). Still, as the authors envisage, there *could* be a trend towards making "gnomic a most intensively studied language and most intriguing reading". □

Richard Grantham is a Professor in the Institute d'Evolution Moléculaire, Université Claude Bernard Lyon 1, 69622 Villeurbanne cedex, France.

Communication between segments of DNA during site-specific recombination

Martin Gellert* & Howard Nash†

* Laboratory of Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

† Laboratory of Molecular Biology, National Institute of Mental Health, Bethesda, Maryland 20892, USA

Some site-specific recombination systems require the interacting DNA sequences to have a specific relative orientation. This means that DNA segments can sense each other's direction even though they may be separated by many thousands of base pairs. Here, we review the surprising results of recent experiments that lead to a new model which accounts for site orientation specificity and relates it to other recombination systems where relative orientation is not critical.

A GROWING number of recombination reactions in both eukaryotes and prokaryotes are known to entail the interaction of specific DNA sequences. Site-specific recombination underlies the integration of the DNA of many bacterial viruses into their host chromosomes as well as the capacity of some bacterial viruses to vary their host range by expressing alternative tail fibre genes. The assembly of the genes encoding antibodies results from recombination between specific sequences flanking the separate units that make up the complete gene, and drug-resistance elements move from one segment of DNA to another by a related process. Much the best understood of these site-specific recombination reactions are those involving the circular DNA molecules of bacteria and their viruses and transposons. In these cases, biochemical investigation has made rapid progress; a number of site-specific recombination reactions can now be studied in completely defined protein systems. A major challenge in the study of such reactions has been to understand how proteins can find, bring together and recombine two defined segments of DNA with high efficiency and accuracy.

In principle, two targets of recombination can be arranged on DNA circles such as the chromosomes of prokaryotes in any of three ways. If they are on separate DNA molecules, as when the circular form of the bacteriophage λ DNA integrates into a bacterial genome, recombination of the two sites joins the two DNAs to form a single circular molecule (Fig. 1a). If the two sites are on the same DNA molecule then, because the specific sequence of such sites is asymmetric, they can be arranged either facing the same way (colinear, Fig. 1b) or in opposite orientations (inverted, Fig. 1c). Recombination of colinear sites resolves one circle into two (Fig. 1b): this is how the two copies of a drug-resistance element, formed in a single circle in the first stage of transposition, are separated to complete the process. Recombination of inverted sites in the same DNA circle (Fig. 1c) flips one of the regions between the sites in relation to the other: this is the mechanism used in viruses where inverting a gene alters the host range. A surprising finding from studies *in vitro* has been that although some systems carry out recombination with two sites in any arrangement, other systems have a strict requirement for just one arrangement.

For example, resolvase proteins of the drug-resistance element Tn3 and of the related transposon Tn1000 (refs 1, 2), require the sites to be on the same circle and facing the same way. DNA invertase proteins like Hin, which switches the expression of the surface antigens of *Salmonella* and Gin, which alters the host range characteristics of bacteriophage Mu, demand that the sites be on the same molecule but in opposite orientations^{3,4}. Although they do not recombine sites on separate DNA molecules, both DNA resolvases and DNA invertases are rather insensitive to the distance between sites on the same molecule; sites separated by a few hundred or many thousands of bases

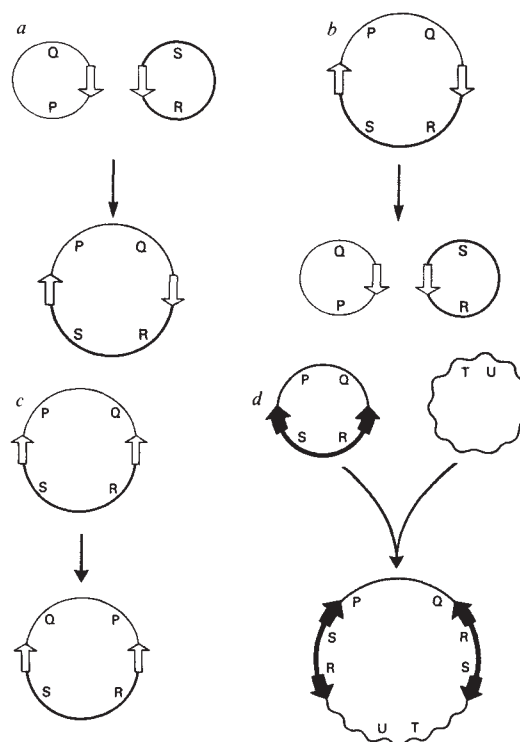


Fig. 1 Arrangement of DNA sites in various recombination reactions. Double-helical DNA is shown as a single line; the two halves of the substrate that are rearranged by recombination are distinguished by lines of different thickness. Letters denote various genetic regions. Recombination sites are shown as arrows to indicate that each site has a polarity. For the recombination reactions shown in a, b and c, recombination can be pictured as the result of breakage of DNA within the arrows followed by rejoining of the broken segments to form recombinants. Depending on the location and relative orientation of the sites, this leads to the union of two circles (a), the resolution of one circle into two (b), or the inversion of one segment of a circle with respect to the other (c). In d, a transposon (solid arrows and connecting heavy line segment) is shown as part of a DNA circle. Transposition involves the replication of the transposon coupled to the joining (at both arrowheads) to a random site on a separate DNA molecule (wavy line). For transposition of bacteriophage Mu, the resulting cointegrate appears to be a stable final product. For transposition of elements such as Tn3 and Tn1000, the cointegrate can be a substrate for a further reaction. Although not shown in d, these two transposons contain, in addition to their recombinogenic ends, sites like those shown in a-c; in a cointegrate these sites are arranged as in b and are targets for a DNA resolvase that is encoded by the transposon.

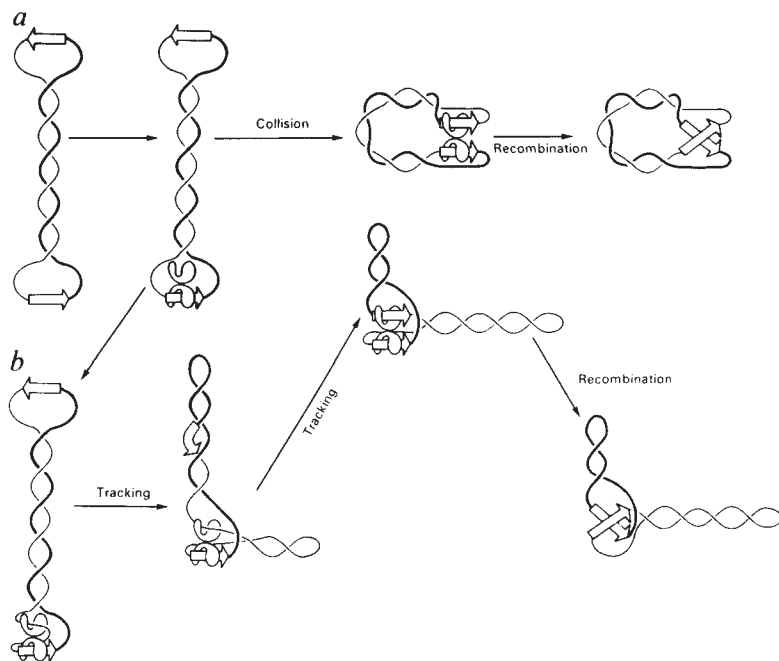


Fig. 2 Pathways for juxtaposition of recombining sites. A supercoiled double-helical circle is shown as an interwound single line. Within this circular substrate the colinear recombining sites are designated by open arrows, the two substrate regions by lines of different thickness and the recombination apparatus by kidney-shaped structures. In *a*, the sites come together by random collision and recombination produces a highly interwound catenane. In *b*, the DNA is threaded through the recombination apparatus to bring the sites together in an ordered manner so that recombination generates a catenane with only one link.

are handled with comparable efficiency. Orientation specificity is the rule again in the transpositional recombination of bacteriophage Mu, which proliferates by the process illustrated in Fig. 1*d*. This form of transposition involves the joining of two DNA circles; one (the donor) contains the transposon, and the other is a non-specific target DNA. Coupled to this joining process, there is an obligatory replication of the transposon DNA to form a 'cointegrate' which combines donor and target circles. For successful transposition, the ends of the Mu transposon must be located on the same donor DNA and must be oriented inversely^{5,6}.

The requirement for connected sites that have a specific relative orientation implies communication between DNA segments. How are specific recombination proteins able to distinguish the orientation of the recombining sites? To put the problem in context, first consider reactions where orientation effects are not important. The integration of phage λ is one example. As it occurs in nature, this reaction recombines a site on the phage DNA (*attP*) with a site on the host DNA (*attB*). But *attP* and *attB* can be artificially placed in a single circular DNA molecule and recombination can then be shown to be catalysed efficiently with sites facing in either the same or opposite directions⁷. The DNA must however be negatively supercoiled for recombination to occur and this, together with the relative orientation of recombining sites, strongly influences the nature of the recombination products. When the two sites are colinear, recombination produces catenated circles, whereas if the sites are inverted, the event gives rise to a single knotted molecule. In either case, the topological complexity of the product rises as the DNA supercoiling increases^{8,9}. One can explain the results by supposing¹⁰ that the two sites in an interwound superhelix encounter each other at random and trap a variable number of superhelical turns between them (Fig. 2*a*). This hypothesis is strongly supported by the topological structure of the resulting catenanes and knots¹¹. Also in keeping with the idea that attachment sites come together by random collision is the finding that recombination between *attP* and *attB* *in vitro*, as *in vivo*, is efficient even when the sites are on separate pieces of DNA¹². Other cases in which recombination is insensitive to the orientation of the target sites, are, for example, the reactions promoted by the Cre protein of bacteriophage P1 (ref. 13) and the rearrangement of variable and joining segments of immunoglobulin genes¹⁴.

What is different about the more fastidious reactions where

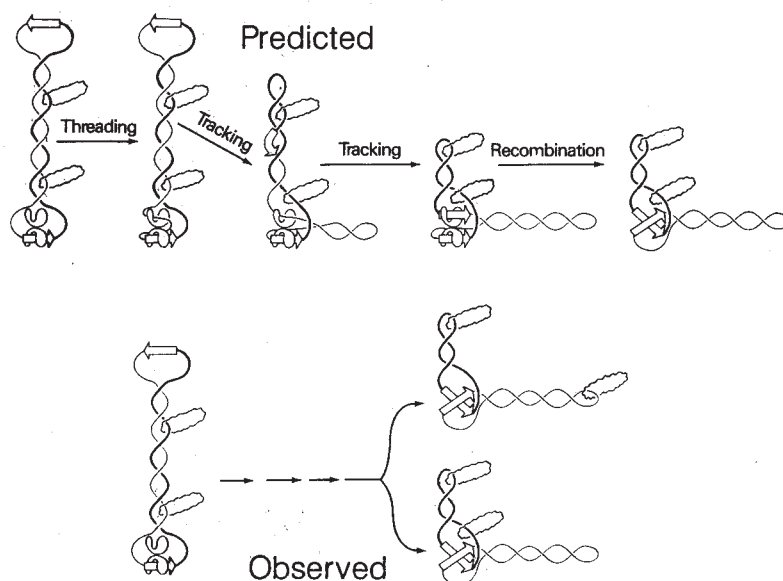
correct site orientation is crucial? We shall show here how recent experiments have ruled out one class of special mechanisms as the basis of orientation sensitivity, and suggest how all site-specific recombination systems may use a common principle to bring sites together. Some of these experiments have already been briefly touched on in a recent review in *Nature*¹⁵ discussing how the same principle can be applied to transcriptional activation.

Tracking derailed

One possible explanation for the orientation sensitivity of some reactions has become known as protein tracking^{2,16,17} or sliding¹⁵. In its simplest version (Fig. 2*b*), it supposes that a recombination protein, bifunctional or dimeric in structure, uses one domain to recognize and attach tightly to one of the recognition sites on the DNA. The other domain of the recombinase binds non-specifically to DNA near this site; the loosely bound DNA is postulated to remain in contact with the recombinase but to be capable of sliding along it until a second site is encountered. If this site is in the correct orientation, the protein catalyses the recombinational strand exchange. If the second site is incorrectly oriented, it will not be recognized; if it is on a different DNA chain the protein will never see it, having both domains occupied by a single DNA molecule. This picture, although attractive, is not free of problems. One difficulty is that many of the recombination systems cited above do not require an energy-donating cofactor; protein tracking thus cannot be driven in one direction, but must occur by diffusion along the DNA. One might then expect the capture of the second site to be time-consuming when the distance between sites is large, though it is hard to put a quantitative bound on the rate.

Several lines of experimentation have nonetheless been interpreted as favouring a tracking mechanism. When the Tn3 resolvase is allowed to act on a plasmid containing four recombination (*res*) sites, all in the same orientation so that all pairs are able to recombine, neighbouring sites are preferred to non-adjacent sites as partners¹⁸. Clearly, if a protein started at one site and searched for others without releasing its grip on the DNA, it would encounter neighbouring sites first. The topology of molecules recombined by Tn3 resolvase has also been interpreted as being consistent with a tracking mechanism¹⁸. Irrespective of the degree of supercoiling in the substrate, the two

Fig. 3 Reporter ring experiment. At the left, a recombination substrate like that in Fig. 2 is shown catenated to two unreactive DNA circles (wavy loops). In the upper panel, successive drawings show the progress of a tracking mechanism for juxtaposition of recombining sites (open arrows). The predicted result is a structure in which both reporter rings are linked to only one of the two recombinant circles. In the lower panel is shown the observed result. In contrast to the prediction of the tracking model, half of the recombinants have one reporter ring on each of the two product circles.



recombinant DNA rings are found in catenanes that contain just one link (Fig. 2b). This would not happen if two sites in different parts of a supercoiled DNA met at random (as seems to be true in the λ integration reaction described above).

However, a more subtle experiment designed to test the tracking model has given an unexpected answer that casts doubt on this interpretation. In this experiment¹⁸, a substrate with two colinear Tn3 *res* sites was catenated to several circles of an unreactive DNA before undergoing recombination (Fig. 3). The expectation was that, as substrate DNA was threaded through resolvase, the unreactive DNA circles (reporter rings) would be left behind and thus excluded, in each event, from one of the two recombinant product circles. Tracking could start from either site, so it would not always be the same recombinant circle from which reporter rings were excluded; but in no case should the distribution of rings be random. In fact, an essentially random distribution was found. Although there are ways of modifying tracking models to make them compatible with this result (for instance by allowing the protein to hop over obstructions while maintaining its path along the chosen helix), it has drastically reduced the appeal of such explanations.

Two more recent lines of experimentation from the laboratories of Sherratt and Mizuuchi have now convincingly eliminated tracking models; instead they suggest an altogether different mechanism. Boocock *et al.*¹⁹ begin by finding reaction conditions in which Tn3 resolvase no longer requires a supercoiled DNA substrate, so that recombination may occur with nicked circular DNA or linear DNA. Quite remarkably, although the requirement for sites to face the same way is maintained when nicked circular DNA is used in place of supercoiled DNA, it disappears when linear DNA is used. With linear substrates, inversion is permitted and intermolecular recombination, an undetectable reaction with circular substrates, is readily observed. It is hard to explain how the tracking model could accommodate the conclusion that strict orientation preference depends on the absence of a free end. An even more direct refutation of the tracking model comes from ingenious experiments on bacteriophage Mu transposition²⁰. In this work a plasmid was constructed in which the Mu ends, correctly oriented for recombination, are separated on one side by a λ *attP* site and on the other by a λ *attB* site, the λ sites being arranged colinearly (Fig. 4). Int-promoted recombination within the supercoiled DNA now divides the substrate into two circles and produces a catenane, with one Mu end residing in each of the interlinked rings. It is thus impossible for a protein to start at one Mu end and continually track along one DNA chain to

reach the other Mu end. Nonetheless, the catenated DNA is found to be active in Mu transposition. This is in sharp contrast to the lack of recombination seen when the Mu ends are on unlinked circles. What about the possibility that the Mu transposase makes short hops from one DNA to another in response to the complexity of catenation? To test this, Craigie and Mizuuchi built a plasmid in which directly repeated (and thus incorrectly oriented) Mu ends were separated by *attP* and *attB*. The catenane derived from this structure by λ recombination failed to transpose Mu. Other experiments, in which recombination between *attP* and *attB* was used to generate knotted molecules containing a pair of Mu ends, provided two more results where tracking fails to account for the effectiveness of recombination substrates²⁰.

The topological filter

Because orientation dependence in Tn3 resolution and Mu transposition cannot be explained by tracking, we must look elsewhere for an acceptable mechanism. Other processive models, such as those that invoke the construction of a protein scaffold that initiates at one site and extends (or oozes¹⁵) along the DNA until it reaches a second site, are eliminated by the same experiments that rule out the tracking model. However, a successful alternative to tracking and scaffolding has been suggested in somewhat different forms by both Boocock *et al.*¹⁹ and

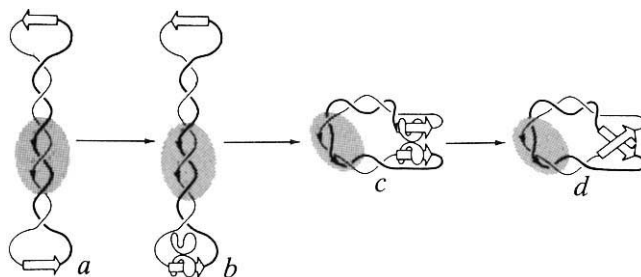


Fig. 4 Simple and catenated substrates for Mu transposition. An inversely oriented set of Mu ends is depicted as a pair of filled arrowheads; stippling highlights the region of the substrate that contains these sites. The circular DNA containing this transposon also carries a pair of λ attachment sites (open arrows) which mark off regions shown as lines of differing thickness. Integrative recombination between the λ sites takes place as described in Fig. 2a, producing a catenane in which each Mu end is on one of a pair of interlocked circles.

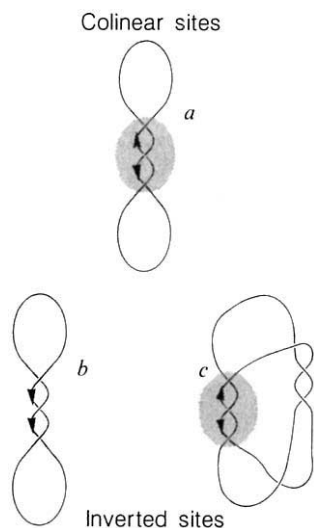


Fig. 5 Synaptic topology with circular substrates. Recombination sites are denoted by filled arrowheads arranged as a direct repeat (a) or an inverted repeat (b, c). In each case, the sites are pictured to have become interwound in a way that produces three nodes, places where the DNA helix crosses itself. Each node has a negative topological sign²³, the same as that of the nodes in negatively supercoiled DNA. Stippling outlines the arrangement that is presumed to be recognized by the recombination apparatus; heavy and light stippling imply that enzymes with different specificity must recognize the two different arrangements of sites. For example, while resolvase has been postulated¹⁹ to recognize the arrangement shown in a, systems like Mu transposition that require inverted sites should recognize (as shown in Fig. 4 and ref. 20) the arrangement shown in b. In c, two inverted sites are interwound so as to produce the same configuration as that shown in a. To achieve this geometry the remaining DNA must form a compensating tangle.

Craigie and Mizuuchi²⁰. The basic idea is that recognition of site orientation reflects constraints that derive, first, from the circularity of the DNA substrate and, second, from a requirement for highly specific interwrapping of the recombining sites. The combined effect of these two constraints has been described as a 'topological filter' that limits productive synapsis to a particular arrangement of sites²¹. The concept of interwrapping of recombining sites has found experimental support in the work of Cozzarelli and co-workers who showed that the topology of the recombinants produced by resolvase could be best explained if, at the moment of crossing over, the two halves of the substrate DNA writhe about one another three times^{22,23}. Recently, cross-linking experiments have provided additional evidence that interwrapping of sites occurs during synapsis (H. Benjamin and N. Cozzarelli, personal communication). Boocock *et al.*¹⁹ and Craigie and Mizuuchi²⁰ point out that supercoiling will favour one set of such interwrappings but inhibit others. The tendency of supercoiled DNA to relieve its strain should promote the interwrapping that is required for the resolvase synapse. When

the sites are colinear, the structure that forms is presumed to be primed for recombination (Fig. 5a) but when the sites are inversely repeated, supercoiling will favour a structure with a different (and inappropriate) alignment of sites (Fig. 5b). To form the 'correct' synaptic structure with a substrate containing inverted sites, one must make a tangled circle (Fig. 5c). The same complication arises during synapsis of sites on separate circles: interwrapping of sites is accompanied by tangling. Intuitively, such an elaborate tangle seems implausible and, if it is energetically costly, synapsis of the inappropriately oriented sites should be inhibited. Moreover it is then easy to see why orientation dependence is not important when the recombination substrate is made linear¹⁹; the free ends permit the DNA to escape permanent entanglement. The behaviour of the catenated and knotted substrates in the experiments of Craigie and Mizuuchi²⁰ can also be explained in this framework. Assuming that, before Int-promoted recombination, the structure of the circular substrate favours the correct synaptic geometry (Fig. 4a), then this configuration will not be affected when the connecting DNA outside the synaptic region is rearranged by integrative recombination (Fig. 4c and d). It is the local geometry of the two sites that is critical, not the presence or absence of a continuous DNA path connecting them.

Coupled with the experiments described above, the emergence of a plausible model for site orientation specificity based on interwrapping of sites virtually completes the rejection of mechanisms that depend on processive measures such as protein tracking. What is particularly satisfying about this new view of dependence on site orientation is the way it provides a unified understanding of several apparently different systems. The lack of orientation specificity for λ integrative recombination suggested that the pathway by which λ attachment sites find each other would be very different from that used by resolvase and transposase. It now seems that all three systems use a common mechanism to bring sites together. For all three cases, sites are juxtaposed by random collision; in the λ system all collisions can lead to recombination, but in the other systems a requirement for interwrapping of the sites can place restrictions on which encounters are productive. An important challenge for future research is to understand the role of interwrapping in recombination and to determine the degree to which each recombination system depends on it.

The insights gained in the work described above can be compared with those arising from studies of the regulation of gene expression. There, it appears that proteins bound to enhancers, upstream activators, operators and promoters can also interact with distant regulatory sites by simple collision^{15,24,25}. Thus in both recombination and transcription systems, direct interaction between proteins bound to DNA may be a governing principle. With the advent of purified systems for a wide range of transcriptional processes, it will be interesting to see whether topological methods can be used to devise tests of mechanisms proposed for the control of gene expression.

We thank Mark Ptashne for encouraging us to write this article. We are indebted to him and to our many colleagues at NIH for suggesting improvements in the manuscript. We also acknowledge the cheerful and effective help of Joanne Russell in preparing the figures.

1. Reed, R. R. *Cell* **25**, 713-719 (1981).
2. Kitts, P. A., Symington, L. S., Dyson, P. & Sherratt, D. J. *EMBO J.* **2**, 1055-1060 (1983).
3. Plasterk, R. H. A., Ilmer, T. A. M. & van de Putte, P. *Virology* **127**, 24-36 (1983).
4. Johnson, R. C., Bruist, M. F., Glaccum, M. B. & Simon, M. I. *Cold Spring Harbor Symp. quant. Biol.* **49**, 751-760 (1984).
5. Howe, M. M. & Schumm, J. W. *Cold Spring Harb. Symp. quant. Biol.* **45**, 337-346 (1981).
6. Mizuuchi, K. *Cell* **35**, 785-794 (1983).
7. Mizuuchi, K., Fisher, L. M., O'Dea, M. H. & Gellert, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1847-1851 (1980).
8. Pollock, T. J. & Nash, H. A. *J. molec. Biol.* **170**, 1-18 (1983).
9. Bliska, J. & Cozzarelli, N. *J. molec. Biol.* (in the press).
10. Mizuuchi, K., Gellert, M., Weisberg, R. A. & Nash, H. A. *J. molec. Biol.* **141**, 485-494 (1980).
11. Spengler, S. J., Stasiak, A. & Cozzarelli, N. R. *Cell* **42**, 325-334 (1985).
12. Mizuuchi, K. & Mizuuchi, M. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1111-1114 (1979).
13. Abremski, K., Hoess, R. & Sternberg, N. *Cell* **32**, 1301-1311 (1983).

14. Lewis, S., Gifford, A. & Baltimore, D. *Science* **228**, 677-685 (1985).
15. Ptashne, M. *Nature* **322**, 697-701 (1986).
16. Krasnow, M. A. & Cozzarelli, N. R. *Cell* **32**, 1313-1324 (1983).
17. Craig, N. L. & Nash, H. A. in *Mechanisms of DNA Replication and Recombination* (ed. Cozzarelli, N. R.) 617-636 (Liss, New York, 1983).
18. Benjamin, H. W., Matzuk, M. M., Krasnow, M. A. & Cozzarelli, N. R. *Cell* **40**, 147-158 (1985).
19. Boocock, M. R., Brown, J. L. & Sherratt, D. J. *Biochem. Soc. Trans.* **14**, 214-216 (1985).
20. Craigie, R. & Mizuuchi, K. *Cell* **45**, 795-800 (1986).
21. Boocock, M. R., Brown, J. L. & Sherratt, D. J. in *DNA Replication and Recombination* (eds Kelly, T. J., Jr. & McMacken, R.) (Liss, New York, in the press).
22. Wasserman, S. A. & Cozzarelli, N. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1079-1083 (1985).
23. Wasserman, S. A., Dungan, J. M. & Cozzarelli, N. R. *Science* **229**, 171-174 (1985).
24. Dynan, W. S. & Tjian, R. *Nature* **316**, 774-778 (1985).
25. Echols, H. *Science* **233**, 1050-1056 (1986).

Three-dimensional structure of the Earth from splitting in free-oscillation spectra

Domenico Giardini, Xiang-Dong Li & John H. Woodhouse

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, 02138, USA

Observations of splitting in free oscillations spectra can be used to constrain the three-dimensional distribution of heterogeneity in the Earth. Modes sampling mainly the Earth's mantle are compatible with the distributions of mantle heterogeneity inferred from P-wave travel times. The anomalously large splitting of modes penetrating deeply into the core requires large departures from sphericity which are difficult to accommodate within the framework of an isotropic model; it is argued that anisotropy of the inner core offers the most natural explanation of these data.

FOLLOWING moderately large earthquakes seismic waves are observed worldwide; the recordings of these disturbances constitute the primary data source for studies of the Earth's internal structure. Under the hypothesis of isotropy, the physical parameters to which such observations are sensitive are the bulk modulus $\kappa(r, \theta, \varphi)$, the shear modulus $\mu(r, \theta, \varphi)$, and density $\rho(r, \theta, \varphi)$, where r, θ, φ are radius, colatitude and longitude. To accommodate dissipative effects κ and μ may be thought of as possessing small imaginary parts: $\kappa = \kappa_0(1 + iQ_\kappa^{-1})$, $\mu = \mu_0(1 + iQ_\mu^{-1})$ where $\kappa_0, \mu_0, Q_\kappa, Q_\mu$ are real and have a weak dependence on frequency^{1,2}.

Studies since the beginning of the century have sought to determine the spherically symmetric parts of the parameter distributions $\bar{\rho}_0(r)$, $\bar{\kappa}_0(r)$, $\bar{\mu}_0(r)$, $\bar{Q}_\kappa(r)$, $\bar{Q}_\mu(r)$, resulting in spherical models of the Earth which are close to optimal in predicting a very wide range of seismic observations. Remaining discrepancies between observations and model predictions are largely due to deviations of the Earth from spherical symmetry, and while such deviations are relatively small, they constitute the principal observable evidence of processes taking place in the Earth's interior; as such, they have become a major focus of seismological research.

Here we take the velocities of shear (S) and longitudinal (P) waves: $v_s = [\mu/\rho]^{1/2}$, $v_p = [(\kappa + \frac{4}{3}\mu)/\rho]^{1/2}$ together with the density, ρ , to be the fundamental mechanical parameters, and write $v_s = \bar{v}_s[1 + \varepsilon_s(r, \theta, \varphi)]$, $v_p = \bar{v}_p[1 + \varepsilon_p(r, \theta, \varphi)]$, $\rho = \bar{\rho}[1 + \varepsilon_\rho(r, \theta, \varphi)]$ where $\bar{v}_s, \bar{v}_p$ and $\bar{\rho}$ are taken from a spherically symmetric reference Earth model; $\varepsilon_s, \varepsilon_p$ and ε_ρ provide a dimensionless parameterization of heterogeneity.

The principal subdivisions of the Earth are the crust and upper mantle ($5,701 \leq r \leq 6371$ km), the lower mantle ($3,480 \leq r \leq 5,701$ km), the fluid outer core ($1,221 \leq r \leq 3,480$ km) and the solid inner core ($0 \leq r \leq 1,221$ km). The free surface and the boundaries between these regions are subject to small deflections from sphericity; we write for the radii of these surfaces: $R_{FS} = \bar{R}_{FS}[1 + \delta_{FS}(\theta, \varphi)]$, $R_{CMB} = \bar{R}_{CMB}[1 + \delta_{CMB}(\theta, \varphi)]$, $R_{ICB} = \bar{R}_{ICB}[1 + \delta_{ICB}(\theta, \varphi)]$, where FS, CMB and ICB refer to the free surface, the core-mantle boundary and the inner core boundary.

In recent years there have been a number of studies of large scale heterogeneity in the Earth, based upon several different kinds of seismic data³⁻¹⁶. In particular we refer to models of lower mantle P-velocity based upon travel time residuals^{4,6} and models of upper mantle S-velocity based upon very long period waveforms^{10,16}. We report conclusions based upon observations of spectral splitting of free oscillations, a phenomenon analogous to, for instance, the Zeeman effect in atomic physics, which results from rotation ellipticity and other asphericity of the Earth¹⁷⁻²⁴.

Masters and Gilbert²⁵ have made the surprising observation that the splitting width of the model₁₀S₂ is more than twice that

accounted for by the effects of Coriolis forces and ellipticity. More recently, Ritzwoller and Masters²⁶ and Ritzwoller *et al.*²⁷ have shown that many other multiplets have anomalously large splitting widths, and have concluded that an axisymmetric structural anomaly in the outer core is required to explain the observations, while recognizing that such an anomaly is essentially ruled out by hydrodynamic and thermal considerations (J. Bloxham, personal communication)²⁸. It is the existence of this large effect, signalling an exotic phenomenon in the core, which has provided the motivation for this study.

Here we present the first results of a technique which enables structural information to be derived from seismic spectra; in principle we solve the least-squares inverse problem for structural parameters using the observed spectra as data. Our current data set consists of some 770 spectra for 27 modes—both anomalously and normally split. This data set alone is insufficient to resolve all structural parameters or to construct unique models of the mantle, outer core and inner core. By making use of the information from previous studies, however, we are able to test certain hypotheses and to explore the class of models which are in basic agreement both with previous models of the mantle and with the modal data set. Initially we restrict consideration to isotropic earth models; our conclusion, however, is that the resulting models are sufficiently at variance with other geophysical evidence that the hypothesis of isotropy should be rejected.

We describe two distinct experiments. In the first experiment our aim is to assess the consistency of the modal data with existing mantle models. Taking the subset of 19 modes which are principally sensitive to mantle structure (although many have significant sensitivity to the outer core), we consider models in which v_p, v_s vary proportionally: $\varepsilon_s = \alpha_{SP}\varepsilon_p$, $\varepsilon_\rho = \alpha_{pP}\varepsilon_p$, where α_{SP} and α_{pP} are constants. Results from laboratory experiments on the change in rock properties with temperature²⁹ have sometimes been used^{8,27} as trial values of α_{SP} and α_{pP} : $\alpha_{SP} = 1.25$ and $\alpha_{pP} = 0.5$. The sensitivity of the modal data to mantle heterogeneity in v_s is much greater than its sensitivity to v_p and ρ . Consequently there is only limited resolution for the parameters α_{SP}, α_{pP} ; if, however, we accept the levels of v_p heterogeneity obtained in travel-time analyses^{3,4,6} larger values of α_{SP} are needed: $2 \leq \alpha_{SP} \leq 3$.

Adopting the values $\alpha_{SP} = 2.0$ and $\alpha_{pP} = 0.5$, we invert the mantle mode data set for P-wave velocity in the mantle (for reasons to be discussed below the inversion is limited to the spherical harmonic expansion coefficients of degrees 0, 2 and 4) and compare the resulting distribution of heterogeneity in P-velocity with the previous models. Since the models show substantial similarity we conclude (1) that the mantle mode data are in basic agreement with the models based upon travel time residuals if $\alpha_{SP} \approx 2.0$, and are in fact sufficient to reproduce these models to a fair degree; (2) that despite significant

sensitivity of the 'mantle' modes to outer core structure, heterogeneity in the outer core is not required.

The mantle model we obtain possesses only small zonal terms and is not able to account for anomalous splitting, whose cause, therefore must be sought in the core. In the second experiment our aim is to quantify the core anomalies which are required. Adopting a mantle model constrained to be close to M84A (ref. 10) in the upper mantle and to L02.56 (ref. 4) in the lower mantle ($\alpha_{SP}=2.0$, $\alpha_{OP}=0.5$) and to be consistent with mantle modes we then seek models of the core (including CMB and ICB) and further small adjustments in the mantle to adequately match the entire modal data set. Including, first, heterogeneity in both outer and inner core we find that some modes can be accommodated by large ($\sim 0.5\%$) heterogeneity in the outer core, but that when the data set is taken as a whole the inclusion of outer core heterogeneity leads to models inconsistent with some modes, particularly ${}_3S_2$, ${}_3S_1$ and ${}_2S_3$. Only when inner core heterogeneity is allowed to approach 5% and boundary perturbations of CMB and ICB are allowed to approach 8 km and 25 km can a consistent model be found. Moreover, with these levels of heterogeneity in the inner core and CMB it proves to be possible to dispense entirely with heterogeneity in the outer core. The conclusion of this experiment is first that a large axisymmetric structure, located in the core, is required; and second, that under the hypothesis of isotropy, the modal data require heterogeneity in the CMB, ICB and inner core of the size stated above.

However, these results conflict with recent measurements of the Earth's nutation periods³⁰ and with evidence from PcP and PKP travel times³¹ and we are led to conclude that there is no acceptable isotropic model capable of explaining the modal observations.

The splitting function

In a spherical Earth model a spheroidal multiplet ${}_nS_l$ is characterized by its angular order l (analogous to orbital angular momentum in atomic physics) and radial order n . Its eigenfrequency ω_l is $(2l+1)$ -fold degenerate, and its singlets are labelled by their azimuthal order m ($-l \leq m \leq l$; analogous to magnetic quantum number).

Rotation, ellipticity, or other structural asphericity removes this degeneracy, leading to singlet eigenfrequencies $\omega_l + \delta\omega_l^i$ ($i=1, 2, \dots, 2l+1$). The frequency perturbations $\delta\omega_l^i$ may be calculated as the eigenvalues of the splitting matrix:

$$H_{mm'} = \Omega\beta\delta_{mm'} + \bar{\omega} \sum_{st} c_{st}(2l+1) \frac{(2s+1)^{1/2}}{2\pi^{1/2}} \times (-1)^m \begin{pmatrix} l & l & s \\ -m & m' & t \end{pmatrix} \begin{pmatrix} l & l & s \\ 0 & 0 & 0 \end{pmatrix} \quad (1)$$

where the Wigner 3- j symbol³² has been employed; Ω is the Earth's rotational angular velocity and β is the Coriolis splitting parameter¹⁹. The quantities c_{st} are linearly related to the Earth's heterogeneity of spherical harmonic order s and degree t ($-s \leq t \leq s$):

$$c_{st} = \delta_{s2}\delta_{t0}c^e + \int_0^a \{E_P^s(r)\varepsilon_P^{st}(r) + E_S^s(r)\varepsilon_S^{st}(r) + E_\rho^s(r)\varepsilon_\rho^{st}(r)\} dr + \sum_h D_h^s \delta_h^{st} \quad (2)$$

where the first term represents the theoretical contribution of ellipticity, and $\varepsilon_P^{st}(r)$, $\varepsilon_S^{st}(r)$, $\varepsilon_\rho^{st}(r)$, δ_h^{st} ($h=FS, CMB, ICB$) are the spherical harmonic expansion coefficients of ε_P , ε_S , ε_ρ and δ_h , excluding the contribution of ellipticity. In (2) a is the radius of the Earth and the kernels $E_P^s(r)$, $E_S^s(r)$, $E_\rho^s(r)$ and D_h^s are given by known expressions, in terms of the eigenfunction of a given multiplet³.

A component of elastic displacement corresponding to a particular isolated multiplet can be written as a function of time³³:

$$u(t) = \text{Re}[\exp(i\bar{\omega}t) \cdot \mathbf{r} \cdot \exp(i\mathbf{H}t) \cdot \mathbf{s}] \quad (3)$$

where $\mathbf{r}$ and $\mathbf{s}$ are $(2l+1)$ -dimensional vectors, characterizing the receiver and the source, and $\mathbf{H}$ is the $(2l+1) \times (2l+1)$ splitting matrix of equation (1).

Equations (1) and (3) give the (nonlinear) relation between an observed seismogram and the coefficients c_{st} , and equation (2) gives the (linear) relation between these coefficients and the model parameters. Because of selection rules satisfied by the 3- j symbols in (1), the summation over s contains only terms of even s ; $s=0, 2, 4, \dots, 2l$ (refs 34, 35).

The quantities c_{st} ($s=0, 2, \dots, 2l$; $-s \leq t \leq s$) may be thought of as the spherical harmonic expansion coefficients of a certain function, which is related to the local heterogeneity through the radially dependent kernels in (2). For a given multiplet we define:

$$\gamma(\theta, \varphi) = \sum_{s=0}^{2l} \sum_{t=-s}^s c_{st} Y_s^t(\theta, \varphi)$$

which we term the *splitting function*³⁶ of the multiplet, since knowledge of the splitting function is equivalent to knowledge of the splitting matrix $\mathbf{H}$.

Using (1) and (3) it is straightforward to formulate the discrete, nonlinear, least squares inverse problem for the splitting function, using as data all available spectra for the multiplet; the problem is linearized and solved iteratively. The resulting coefficients c_{st} constitute essentially model independent constraints on the three dimensional structure of the Earth of degree s and order t (the inferred values c_{st} will depend upon the source parameters and reference model used, but weakly). Using this technique we have retrieved the splitting functions of some 27 multiplets, with varying degrees of resolution (A similar technique has been employed by Ritzwoller *et al.*²⁷ in the estimation of the c_{20} coefficients alone). Difficulties in completely resolving the splitting functions using the current data set have led us to pursue, in parallel, a second technique in which inversion for structural perturbations is carried out directly. Earth models are first constructed by linear inversion, using the derived splitting functions as data, and these models are then adjusted iteratively to achieve agreement with the spectra for all multiplets.

Data and inversion results

For a study of this kind we require very long period seismograms produced by large earthquakes and recorded digitally. The limiting factors are that such data have been collected only since 1976 and for rather few stations, and that large earthquakes are rare. The principal source is the International Deployment of Accelerometers (IDA) network³⁷, which provided all the data included in this study. All available traces for the ten largest events since 1977 were employed.

In Fig. 1 are shown some examples of data and theoretical predictions in three spectral windows, each containing one or two multiplets. In Fig. 1a-c the theoretical spectra include only the effects of rotation and ellipticity while in Figs. 1d-f the theoretical predictions are based on splitting functions obtained by inversion. It will be noted that, owing to attenuation and finite record length, it is not possible to resolve the singlets within a multiplet in individual spectra. The power of the splitting function inversion technique is that it allows all spectra for a given multiplet to be simultaneously brought to bear in unmasking the underlying fine structure of the spectrum. Both amplitude and phase are very poorly predicted by rotation and ellipticity alone (Fig. 1a-c), often giving a variance ratio (squared misfit/squared data) greater than 1. For the three spectral windows shown the inversion results are based on 19, 57 and 30 traces respectively, and for all data the (initial, final) variance ratios are (1.79, 0.30), (0.69, 0.19) and (0.90, 0.25). Table 1 gives these parameters for all the spectral windows used.

The splitting width of ${}_{13}S_2$ is estimated to be anomalous by a factor of $R=2.4$ (Fig. 1a, d), in agreement with the value

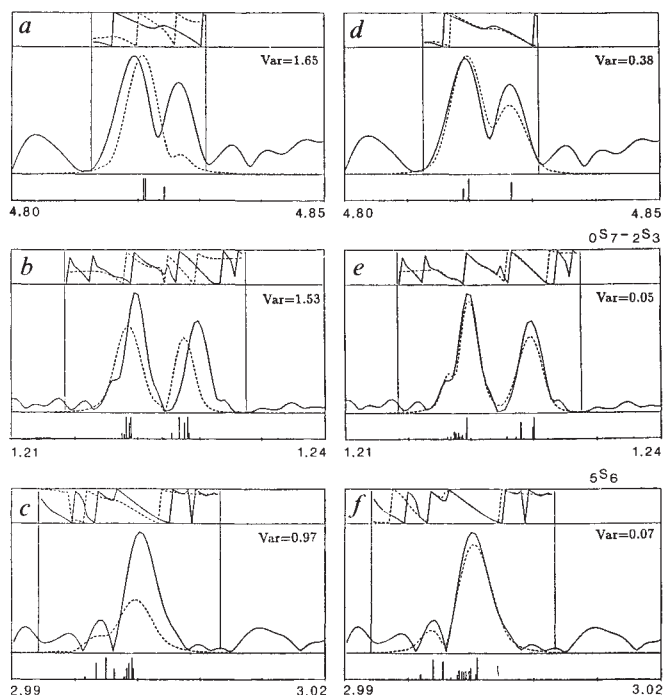


Fig. 1 Examples of data (solid lines) and synthetic spectra (dashed lines) in three spectral windows, each containing one or two multiplets. The range of the (horizontal) frequency axis is indicated in mHz. The spectra are represented in terms of phase in the interval $(-\pi, \pi)$ (top panel of each figure), and amplitude on an arbitrary scale (middle panel). Vertical bars in the bottom panels indicate the frequencies and relative amplitudes of the singlets contributing to the theoretical spectra. The variance ratio (Var) is a measure of the misfit between the observed and synthetic traces shown. In *a-c* the theoretical traces incorporate only the effects of rotation and ellipticity, while in *d-f* the splitting functions found by inversion are employed. A Hann window has been employed to enhance resolution in frequency³⁸.

reported by Ritzwoller *et al.* ($R = 2.3$)²⁷. Since a primary motivation of this study is to find an explanation of such highly anomalous modes as $_{13}S_2$, $_{3}S_2$, which are sensitive only to heterogeneity of degree 0, 2, 4 and since the data set has finite resolution, we limit our inversions both for splitting functions and for model parameters to these degrees, despite the fact that modes of angular order greater than 2 are sensitive to higher degrees; for similar reasons we allow only for degree 0 in the imaginary parts of the splitting functions—corresponding to a spherically symmetric correction in attenuation.

The source parameters used in calculating synthetic spectra are those determined by the centroid-moment tensor technique^{39,40}, subject to corrections in scalar moment using the modal data set; these multiplicative corrections are close to unity and play only a minor role in reducing the data variance; similarly attenuation corrections give small contributions. The large discrepancies in the amplitude spectra (Fig. 1) are accommodated, instead, by adjustment of the interference between closely spaced singlets.

The modes used in this study were selected to span a wide range in phase velocities and to be easily observable in many individual spectra. Notable modes which have been omitted here are the closely spaced pair $_{10}S_2$, $_{11}S_2$. These modes have eigenfunctions which are poorly predicted by current spherical models²⁵. The observed mode can be obtained by taking a linear combination of eigenfunctions predicted by PREM⁴, resulting in a mode very similar in its properties and its splitting to $_{13}S_2$

Table 1 Normal modes and variance ratios from different models

Mode	T	V_0	V_{Cst}	V_m	Rec
$0S_3^+$	2135	0.206	0.169	0.202	22
$1S_3^+ - 3S_1^+$	1061	0.436	0.302	0.456	37
$0S_6^+$	964	0.624	0.144	0.170	37
$3S_2^+$	904	1.467	0.323	0.401	25
$1S_4^+$	851	0.400	0.325	0.329	35
$0S_7^+ - 2S_3^+$	809	0.689	0.188	0.169	57
$1S_5^+ - 2S_4^+$	727	0.659	0.341	0.543	47
$2S_3^+ - 1S_6^+$	658	0.876	0.727	0.819	35
$1S_7^+$	604	0.880	0.282	0.359	38
$1S_8^+$	556	0.936	0.311	0.441	36
$4S_3^+ - 2S_8^+$	488	0.862	0.450	0.530	67
$5S_4^+$	461	0.554	0.377	0.472	11
$5S_4^+$	420	0.687	0.402	0.521	47
$5S_8^+$	370	0.708	0.349	0.439	58
$3S_8^+ - 6S_3^+$	354	0.944	0.355	0.494	39
$5S_6^+$	332	0.900	0.252	0.476	30
$9S_3^+$	281	1.020	0.302	0.433	15
$11S_4^+$	210	1.406	0.406	0.539	22
$13S_2^+$	207	1.790	0.304	0.377	19
$11S_5^+$	197	1.140	0.556	0.747	24
$13S_3^+$	193	1.069	0.398	0.421	21

The frequency windows are ordered by decreasing period in seconds (T). The 'mantle modes' are indicated by $+$. V_0 is the variance ratio (squared misfit/squared data) obtained using the reference model PREM⁴² and incorporating only the effects of rotation and ellipticity (Fig. 1*a-c*); V_{Cst} is the variance obtained using the splitting functions retrieved by nonlinear inversion (cf. Fig. 1*d-f*); V_m is from the anisotropic model discussed in the text. The number of records used for each spectral window is indicated.

and having a Q of approximately 800. A more complete analysis of this mode will be presented elsewhere.

Aspherical models

Before describing the results of inversion for model parameters, we first discuss some examples of splitting functions retrieved by the technique outlined before, together with their associated kernels, which quantify the sensitivity of the splitting function to perturbations in ϵ_p , ϵ_s and ϵ_ρ as a function of radius and to δ_{FS} , δ_{CMB} and δ_{ICB} .

The modes $1S_7$, $0S_6$ and $1S_8$ (see Fig. 2*a-c*) are characterized by high sensitivity to S -velocity in the lower mantle. Note that if $\alpha_{SP} = 2.0$, a truer representation of the importance of E_S , relative to E_p , would be obtained by multiplying E_S by this factor. The splitting functions of these modes are similar and are dominated by a degree 2 pattern known to be characteristic of the lower mantle⁴. Notwithstanding significant sensitivity in the outermost core these splitting functions do not possess the large axisymmetric terms characteristic of anomalously split modes, suggesting that this region is not responsible for anomalous splitting.

The multiplets $_{13}S_3$, $6S_3$ and $_{13}S_2$, on the other hand (Fig. 2*d-f*), possess large c_{20} terms (the harmonic $P_2^0(\cos \theta)$). Modes $_{13}S_3$ and $_{13}S_2$ have rather similar kernels, but $_{13}S_2$ shows a larger effect. The fact that the kernels for $_{13}S_2$ are larger in the inner core is suggestive of an inner core origin for the c_{20} anomaly.

The multiplets $3S_1$, $3S_2$ and $2S_3$ (Fig. 2*g-i*) possess sensitivity throughout the Earth and their kernels are, in some respects, similar. These modes have very different splitting functions, however, which constitute rather delicate discriminants between hypotheses concerning the location of the heterogeneity. Mode $3S_1$ (Fig. 2*g*), which is insensitive to degree 4, has a very small splitting function. Its insensitivity to lower mantle structure is probably due to cancellation of P and S effects, since the kernels are of opposite sign and roughly $E_p \sim -2E_s$; the fact that its kernels are large in the outer core makes it difficult to place significant heterogeneity of degree 2 in this region. Mode $3S_2$

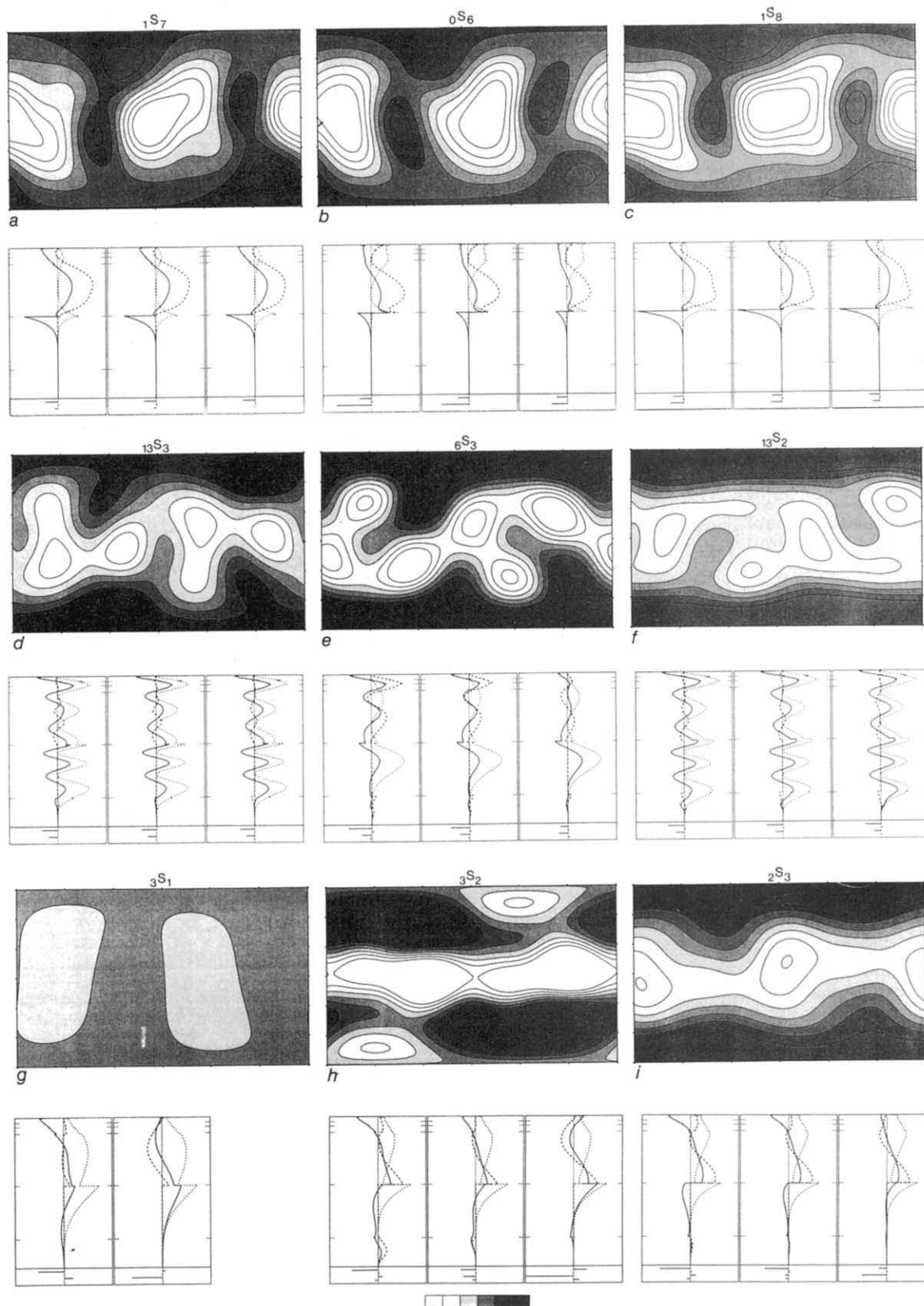


Fig. 2 Examples of splitting functions and differential kernels for 9 of the multiplets used. The kernels E_p^s (solid), E_p^s (dotted) and E_s^s (dashed) indicate, on a uniform scale and as a function of radius, running vertically, the sensitivity of the splitting function to a 1% change in ρ , v_p or v_s . The four horizontal bars in the bottom panel of each figure represent, on the same scale, the sensitivity to a deflection of the boundaries by 100 km. These boundaries, from top to bottom, are the free surface, the 670 km discontinuity, the core-mantle boundary and the inner core boundary. The kernels shown (from left to right) correspond to heterogeneity of harmonic degree 0, 2, 4. The line corresponding to a degree 4 perturbation of CMB in $h(\text{mode } 3S_2)$ exceeds the scale by 18%. The splitting functions are represented in a global linear map projection (ordinate latitude, abscissa longitude) centred on longitude 180°. The contouring interval is 0.067%; the maximum of the scale of shading, which is saturated in some cases, is 0.2%; darker shading corresponds to positive values.

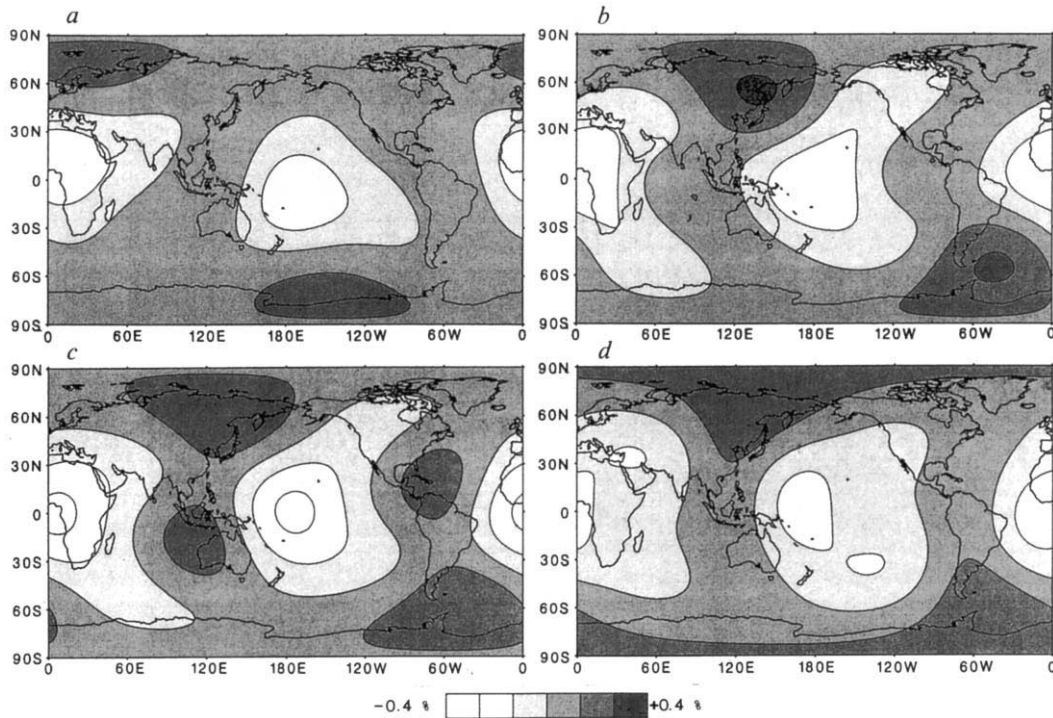


Fig. 3 Comparison of Model 1 (left; *a, c*) with the travel-time model V3.I of Morelli & Dziewonski⁶ (right; *b, d*) at depths 2,700 km (top; *a, b*) and 2,300 km (bottom; *c, d*). Correlation coefficients are $cc_{ab}=0.74$, $cc_{cd}=0.88$. The models correspond to completely independent inversions of very different data sets. The Model V3.I contains all degrees up to $s=6$, but only degrees 2 and 4 are shown.

has, among the multiplets studied, by far the greatest components of degree 4 (Fig. 2*h*). Its great sensitivity to a degree 4 deflection of the core-mantle boundary suggests that the CMB can give a large contribution for this mode. Its kernels in the mantle and in the outermost core are similar to those of ${}_2S_3$ (Fig. 2*i*), which, in contrast, has a splitting function dominated by c_{20} and, to a lesser extent by other terms of degree 2. The correlation coefficients of the degree 2 terms, omitting c_{20} , of the splitting function of ${}_2S_3$ with those of the 'mantle' modes ${}_1S_7$, ${}_0S_6$ and ${}_1S_8$ (Fig. 2*a-c*) are 0.96, 0.83 and 0.97, respectively – a strong indication of a common origin for these terms in the mantle. It should be noted that the inversion for each spectral window is carried out independently of other spectral windows, and thus similarities between the derived splitting functions of different multiplets which sample the Earth in a similar way provides confirmation of the accuracy of the results.

The mantle. Taking as data the splitting functions of 19 mantle modes we first invert for a mantle model (see above) using $\alpha_{SP}=2.0$, $\alpha_{PP}=0.5$. The problem is made discrete by parameterising ε_P^{st} (and hence ε_S^{st} , ε_P^{st}) as combinations of orthogonal polynomials of degrees up to 6 defined on the interval $[R_{CMB}, R_{FS}]$. The data set proves insufficient to completely resolve all 105 parameters and we employ a stochastic inverse, which is equivalent to incorporating artificial *a priori* information that the parameters are small. The trade-off between model size and variance reduction leads us to choose a level of damping corresponding to a model possessing 43 degrees of freedom. This model will be referred to hereafter as Model 1. The v_P anomalies in Model 1 have a similar size (in the sense of the r.m.s.) to the travel time model L02.56 of Dziewonski⁴ and the similar model V3.1 of Morelli and Dziewonski⁶.

Here we compare Model 1 with the model V3.I⁶. At the depth 2,700 km ($r=3,671$ km, Fig. 3*a, b*) the dominant features of the models are in fairly good agreement (correlation coefficient on 13 degrees of freedom $cc=0.74$), while at 2,300 km ($r=$

4,071 km, Fig. 3*c-d*) the agreement is somewhat better ($cc=0.88$). The fact that the travel time model at 2,700 km agrees remarkably in shape with the splitting functions of mantle modes (Figs 3*b, 2b*; $cc=0.82$) and that the Model 1 varies relatively slowly with depth in the lowermost mantle (Fig. 3*a-b*) suggests that lack of resolution, rather than any major inconsistency of the travel time model with the splitting function results, is responsible for the poorer agreement at 2,700 km.

In Fig. 4 Model 1 is compared with two other models at depth 1,300 km ($r=5,071$ km). Fig. 4*c* is taken from a recent model of Woodhouse and Dziewonski¹⁶ based upon mantle waves and long period multiple SH reflected body waves, for the upper 1,500 km of the mantle. Since the model is for S-velocity, it is divided by $\alpha_{SP}=2.0$ in order to make the comparison with P. In this case the correlation coefficients of Model 1 with the two other models are somewhat lower, but still highly significant ($cc=0.56$, $cc=0.63$).

The core. As described previously, we construct a model which is designed to be close to L02.56⁴ and M84A¹⁰ in the mantle and to match the modal data. It is also required to predict the observed geoid and to be in mechanical equilibrium with vanishing deviatoric stress in the outer core. The core mantle boundary has large components in the harmonics $P_2^0(\cos\theta)$, $P_4^0(\cos\theta)$, which are required to explain the splitting function of ${}_3S_2$. The inner core boundary perturbations are dominated by degree 2 and exceed 25 km; heterogeneity within the inner core is dominated by degree 2 and exceeds 6% near the top of the inner core, where ε_P reaches its maximum. The axisymmetric anomaly responsible for anomalous splitting is contained in the CMB, ICB and within the inner core. While a model of this kind is capable of explaining most of the modal observations the existence of large CMB topography of degrees 2 and 4 is inconsistent with recent inferences from geodetic data³⁰ and from travel times³¹.

In Fig. 5 the sensitivities of the splitting function coefficients

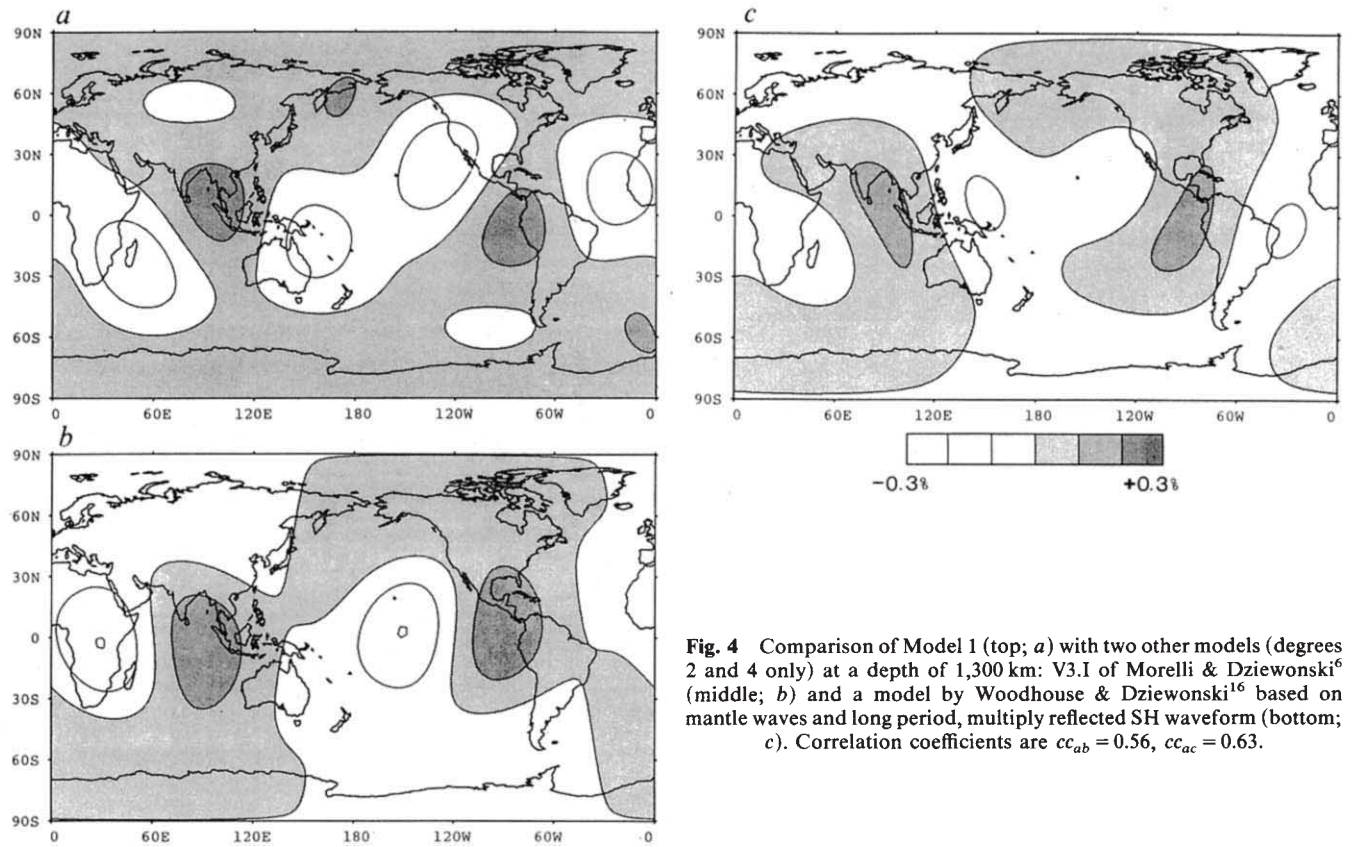


Fig. 4 Comparison of Model 1 (top; *a*) with two other models (degrees 2 and 4 only) at a depth of 1,300 km: V3.1 of Morelli & Dziewonski⁶ (middle; *b*) and a model by Woodhouse & Dziewonski¹⁶ based on mantle waves and long period, multiply reflected SH waveform (bottom; *c*). Correlation coefficients are $cc_{ab} = 0.56$, $cc_{ac} = 0.63$.

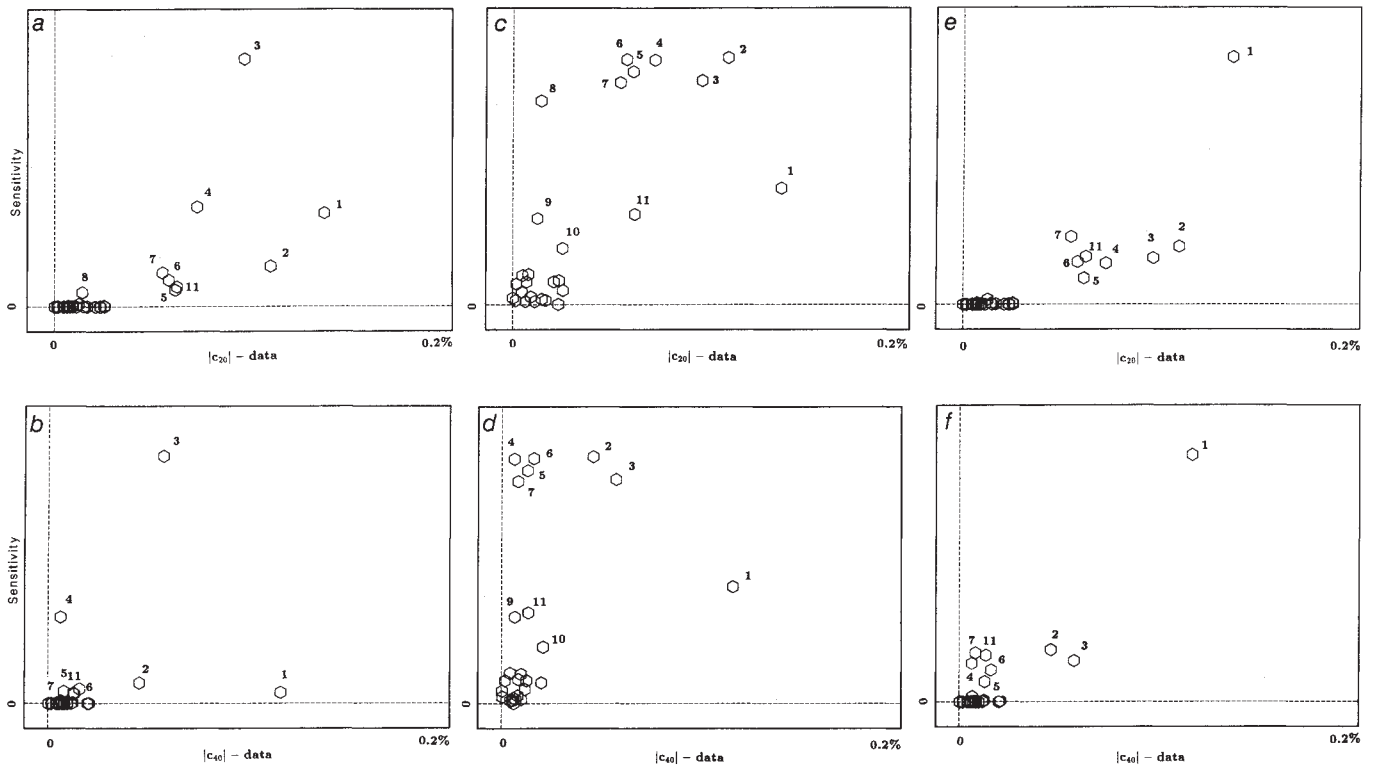


Fig. 5 Plots of the sensitivities of the splitting function coefficients c_{20} (top) and c_{40} (bottom) to structural perturbations in the inner core (left) outer core (middle) and to inner core zonal anisotropy (right). Normal mode multiplets are indicated by numerals: 1: ${}_3S_2$; 2: ${}_6S_3$; 3: ${}_{13}S_2$; 4: ${}_{13}S_3$; 5: ${}_{11}S_5$; 6: ${}_{11}S_4$; 7: ${}_9S_3$; 8: ${}_3S_1$; 9: ${}_5S_3$; 10: ${}_2S_4$; 11: ${}_2S_3$.

c_{20} , c_{40} to structural perturbations in the inner core (Fig. 5a, b); outer core (c, d) and to inner core anisotropy (e, f) are plotted against the observed magnitudes of these coefficients. While it is not possible to draw rigorous conclusions from these, they do provide useful indicators of what kinds of structure are likely to provide plausible explanations of anomalous splitting, and serve as a focal point for discussion.

In Fig. 5a, b, the ordinate is the root mean square sensitivity to inner core ϵ_S and ϵ_P perturbations of degree 2 and 4 respectively; a radial dependence proportional to r^2 is assumed. Modes which do not penetrate the inner core have almost vanishing sensitivities and small observed coefficients. Modes which lie near one of the axes but far from the other represent potential problems for an inner core explanation of anomalous splitting. The mode ${}_3S_2$, for instance, has very small sensitivity to inner core structure but a very large coefficient c_{40} (Fig. 5b). In the model discussed above, this mode could be explained only by CMB topography. The modes ${}_6S_3$ and ${}_{13}S_2$ have coefficients of similar size for both degrees, but very different sensitivities. For this reason they can both be accommodated in the model only by complex radial dependence and by cancellation among ϵ_P , ϵ_S and ϵ_ρ effects for ${}_{13}S_2$.

Figure 5c, d shows similar plots corresponding to an outer core ϵ_P anomaly independent of radius. As noted above, ${}_3S_1$ has a large sensitivity to such a structure and a very small splitting function. The model proposed by Ritzwoller *et al.*²⁷ accommodates this mode by placing a degree 2 anomaly in the mantle which cancels the core effect; it is our conclusion, however, that such an anomaly is inconsistent with the constraints placed on mantle structure by travel time studies⁴⁻⁶ and by mantle modes. For degree 4 (Fig. 5d), modes ${}_{13}S_3$, ${}_{11}S_5$, ${}_{11}S_4$ and ${}_9S_3$ indicate very small heterogeneity but ${}_3S_2$ requires a large effect. An outer core anomaly would need to possess a complicated radial dependence to reconcile the data for degree 4.

Finally, Fig. 5e, f shows sensitivities to axially symmetric anisotropy in the inner core. There is a large number of anisotropic parameters and again the rms sensitivity is used (the details of this calculation will be presented elsewhere). It will be noted that Fig. 5e, f shows a much more coherent pattern. In particular, the highly anomalous mode ${}_3S_2$ has the greatest sensitivity to inner core anisotropy and modes having small degree 4 coefficients also have small sensitivities. These figures show that inner core anisotropy would provide a rather natural

explanation of the character of the anomalous split modes. Table 1 gives the data variances of a model constrained to be close to M84A (ref. 10) and LQ2.56 (ref. 4) in the mantle ($\alpha_{SP} = 2.0$, $\alpha_{PP} = 0.5$) and to have CMB topography close to that determined by Morelli and Dziewonski³¹ and possessing an anisotropic inner core. It will be seen that the single hypothesis of axially symmetric inner core anisotropy yields a rather satisfactory fit to the data. A more complete discussion of anisotropic models will be presented elsewhere.

Conclusions

The splitting functions of free oscillations provide linear constraints on three dimensional Earth models which are (essentially) model independent. Given sufficient data they can be very accurately determined. While the quantity of data now available is marginal for the resolution of some multiplets, the new networks currently under development⁴²⁻⁴⁵, operating over, say, a 10 year span should enable a large number of splitting functions to be as well determined as the geoid, adding greatly to our knowledge of the Earth's internal structure.

The current data set has proved capable of resolving some features of mantle structure already documented using travel-time techniques. In order to satisfy the data for core modes significant zonal heterogeneity in the core is required. Under the hypothesis of isotropy, variations of several percent in inner core properties and undulations of 8 and 23 km on the core-mantle boundary and inner core boundary are required. The alternative hypothesis that the inner core is anisotropic, having an axis of symmetry coincident with the Earth's rotation axis, offers a more natural explanation of the modal observations which, in addition, does not conflict with other data or with the expected homogeneity of the outer core.

We thank J. Berger, D. C. Agnew and the staff of the IDA Project for providing us with the data used in this study. We thank with our colleagues, J. Bloxham, A. M. Dziewonski, A. Morelli and R. J. O'Connell for numerous discussions and we thank A. Morelli and A. M. Dziewonski for allowing us to make use of their unpublished model V3.1. We thank M. Ritzwoller, G. Masters and F. Gilbert for a preprint of their recent paper²⁷ and F. A. Dahlen and F. Gilbert, G. Masters and M. Ritzwoller for their reviews. This research was carried out with the support of the NSF under the grant EAR85-11400.

Received 1 August; accepted 20 November 1986.

1. Futterman, W. I. *J. geophys. Res.* **67**, 5279-5291 (1962).
2. Liu, H.-P., Anderson, D. L. & Kanamori, H. *Geophys. J. R. astr. Soc.* **47**, 41-58 (1976).
3. Dziewonski, A. M., Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **82**, 239-255 (1977).
4. Dziewonski, A. M. *J. geophys. Res.* **89**, 5929-5952 (1984).
5. Clayton, R. W. & Comer, R. P. *Eos* **64**, 776 (1983).
6. Morelli, A. & Dziewonski, A. M. *Eos* **67**, 311 (1986).
7. Creager, K. C. & Jordan, T. H. *Eos* **67**, 311 (1986).
8. Masters, G., Jordan, T. H., Silver, P. G. & Gilbert, F. *Nature* **298**, 609-613 (1982).
9. Woodhouse, J. H. *Proc. Enrico Fermi Int. Sch. Phys.* **85**, 366-397 (1983).
10. Woodhouse, J. H. & Dziewonski, A. M. *J. geophys. Res.* **89**, 5953-5986 (1984).
11. Nakanishi, I. & Anderson, D. L. *Bull. seism. Soc. Am.* **72**, 1185-1194 (1982).
12. Nakanishi, I. & Anderson, D. L. *J. geophys. Res.* **88**, 10267-10283 (1983).
13. Nakanishi, I. & Anderson, D. L. *Geophys. J. R. astr. Soc.* **78**, 573-618 (1984).
14. Nataf, H.-C., Nakanishi, I. & Anderson, D. L. *Geophys. Res. Lett.* **11**, 109 (1984).
15. Tanimoto, T. & Anderson, D. L. *J. geophys. Res.* **90**, 1842 (1985).
16. Woodhouse, J. H. & Dziewonski, A. M. *Eos* **67**, 307 (1986).
17. Ness, N. F., Harrison, C. S. & Slichter, L. J. *J. geophys. Res.* **66**, 621-629 (1961).
18. Benioff, H., Press, F. & Smith, S. W. *J. geophys. Res.* **66**, 605-619 (1961).
19. Backus, G. E. & Gilbert, F. *Proc. natn. Acad. Sci. U.S.A.* **47**, 362-371 (1961).
20. Pekeris, C. L., Alterman, A. & Jarosch, M. *Phys. Rev.* **122**, 1692-1700 (1961).
21. Dahlen, F. A. *Geophys. J. R. astr. Soc.* **16**, 329-367 (1968).
22. Dahlen, F. A. *J. geophys. Res.* **81**, 4951-4956 (1976).
23. Woodhouse, J. H. & Dahlen, F. A. *Geophys. J. R. astr. Soc.* **53**, 335-354 (1978).

24. Dahlen, F. A. & Sailor, R. V. *Geophys. J. R. astr. Soc.* **58**, 609-623 (1979).
25. Masters, G. & Gilbert, F. *Geophys. Res. Lett.* **8**, 569-571 (1981).
26. Ritzwoller, M. & Masters, G. *Eos* **65**, 1003 (1984).
27. Ritzwoller, M., Masters, G. & Gilbert, F. *J. geophys. Res.* **91**, 10203-10228 (1986).
28. Stevenson, D. *Geophys. J. R. astr. Soc.* (submitted).
29. Anderson, O. L., Schreiber, E., Lieberman, R. C. & Soga, M. *Rev. Geophys. Space Phys.* **6**, 491-524 (1968).
30. Gwinn, C. R., Herring, T. A. & Shapiro, I. I. *J. geophys. Res.* **91**, 4755-4765 (1986).
31. Morelli, A. & Dziewonski, A. M. *Nature* (in the press).
32. Edmonds, A. R. *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, 1960).
33. Woodhouse, J. H. & Gernius, T. P. *Geophys. J. R. astr. Soc.* **68**, 653-673 (1982).
34. Zharkov, B. N. & Lyubimov, V. M. *Izv. Acad. Sci. USSR Phys. Solid Earth* **10**, 613-618 (1970).
35. Madariaga, R. thesis, MIT (1971).
36. Woodhouse, J. H. & Giardini, D. *Eos Trans. AGU* **66**, 300 (1985).
37. Agnew, D. C. *et al.* *Eos* **67**, 203-212 (1986).
38. Dahlen, F. A. *Geophys. J. R. astr. Soc.* **69**, 537-549 (1982).
39. Dziewonski, A. M., Chou, T.-A. & Woodhouse, J. H. *J. geophys. Res.* **86**, 2825-2852 (1981).
40. Dziewonski, A. M. & Woodhouse, J. H. *J. geophys. Res.* **88**, 3247-3271 (1983).
41. Dziewonski, A. M. & Anderson, D. L. *Phys. Earth Planet. Inter.* **25**, 297-356 (1981).
42. Romanowicz, B., Cara, M., Fels, J. F. & Rouland, D. *Eos* **65**, 753 (1984).
43. Nolet, G., Romanowicz, B., Kind, R. & Wielandt, E. *ORFEUS science plan, report, Observ. and Res. Facil. for Eur. Seismol. (ORFEUS)*, Luxembourg (1985).
44. Smith, S. W. *Eos* **67**, 213-219 (1986).
45. Romanowicz, B. & Dziewonski, A. M. *Eos* **67**, 541 (1986).

Deregulated *c-fos* expression interferes with normal bone development in transgenic mice

Ulrich Rüther, Christa Garber, Dymitr Komitowski*, Rolf Müller & Erwin F. Wagner

European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, FRG

* Institute of Experimental Pathology, German Cancer Research Center, D-6900 Heidelberg, FRG

Transgenic mice were generated expressing c-fos genes under the control of the human metallothionein promoter. Although high levels of c-fos messenger RNA were detectable in a variety of tissues, deregulated c-fos expression specifically interfered with normal bone development without inducing malignant tumours.

It is generally accepted that gene regulation is crucial for the control of cellular differentiation and development. One class of genes that may be important in these processes is that of the cellular homologues (*c-onc* genes or proto-oncogenes) of retroviral transforming genes (*v-onc* genes)¹. Recent data showing sequence homology between *c-onc* genes and known growth factor or growth factor receptor genes, support the hypothesis that proto-oncogene products are directly involved in normal cellular proliferation and differentiation¹⁻³. It is well known that viral tumorigenesis in experimental animals is caused by *v-onc* genes¹ and it is very likely that structurally altered *c-onc* genes are important in the induction of human malignancies^{1,4,5}. But the possible importance of proto-oncogenes in the differentiation of a pluripotent or multipotent stem cell to a terminal differentiated cell remains largely obscure and often cannot be satisfactorily studied in cell culture systems. Transgenic mice are a powerful alternative system to test the possible function of *c-onc* genes *in vivo*⁶⁻⁸.

We have chosen to study the function of the *c-fos* proto-oncogene in normal development, differentiation and growth control. The *fos* oncogene has been detected in two mouse sarcoma viruses (MSVs), the FBJ-MSV and the FBR-MSV^{9,10}. Although the two viruses contain structurally different forms of the *fos* gene, they both induce osteosarcoma-like tumours in newborn mice. The reason for this tumour type specificity is not known. The normal function of the *c-fos* gene has been the subject of intense investigation^{9,10}. Present evidence suggests that the *c-fos* gene product exerts different functions in different biological processes. A role in growth control, for example, is suggested by the rapid, transient induction of *c-fos* after cells are stimulated with polypeptide growth factors⁹⁻¹³. Also, *c-fos* appears to be important in cellular differentiation, based on the differentiation-promoting potential of exogenous *c-fos* genes introduced into certain teratocarcinoma cell lines^{14,15} and the tissue and cell type specificity of its expression *in vivo*⁹. High levels of *c-fos* expression have been detected in the fetal membranes, yolk sacs and amnion¹⁶, where expression increases as gestation proceeds, in fetal liver and adult bone marrow¹⁷ and in terminally differentiated blood granulocytes¹⁸. Expression of *c-fos* is dependent on external signals in all cases investigated^{9,19} suggesting that its nuclear gene product may be involved in signal transduction by participating in the activation of other cellular genes. This could also explain the various effects of *c-fos*, as different genes may be affected in different cell types, with different phenotypic consequences.

To study the importance of *c-fos* further and to elucidate the mechanism(s) responsible for the specificity of *fos*-induced tumorigenesis, we have generated transgenic mice expressing deregulated exogenous *c-fos* genes in a variety of tissues.

Generation of transgenic mice

The murine *c-fos* gene under the control of the glucocorticoid- and heavy-metal-inducible human metallothionein promoter²⁰

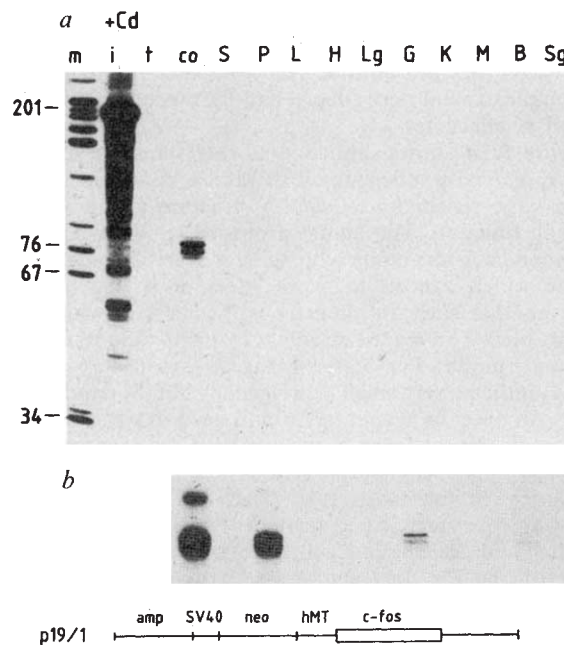


Fig. 1 MT-*c-fos* expression in various tissues of a F₁ offspring from mouse 172-6. RNAs were analysed by an RNase protection assay using a probe which detects the first 75 bases encoded by the hMT mRNA. *a*, Expression analysed after CdCl₂ induction; *b*, expression analysed in a F₁ offspring of 172-6 after cycloheximide (CH) injection. m, Marker; i, product of the sp6 *in vitro* transcripts; t, tRNA; co, control RNA, isolated from a F9 cell clone expressing the MT-*c-fos* gene; S, spleen; P, pancreas; L, liver; H, heart; Lg, lung; G, gonads; K, kidney; M, skeletal muscle; B, brain; Sg, salivary gland. Bottom, injected construct p19/1.

Methods. Siblings of mouse line 172-6 were treated for 6 h before analysis with 100 μ l 10 mM CdCl₂ (*a*) or with CH (two 0.35 ml injections of 5 mg ml⁻¹ in a 3-h interval). RNA was isolated from the individual organs as described⁴¹. RNA (20 μ g) was hybridized overnight at 60 °C with a ³²P-labelled anti-sense *in vitro* transcript, followed by digestion with RNase A and T₁ as described⁴². The protected fragments were separated on a 10% polyacrylamide/urea gel and exposed overnight.

(MT) was introduced into fertilized mouse eggs. This construct¹⁵ (p19/1, Fig. 1) carries the complete *c-fos* gene with its own translation start and stop codons as well as the polyadenylation signal and contains in addition the bacterial *neo* gene with the simian virus 40 (SV40) promoter. For the production of transgenic mice, C3H/HeJ inbred eggs were injected with the linearized construct or the purified MT-*c-fos* fragment without vector sequences, implanted in pseudopregnant females and allowed

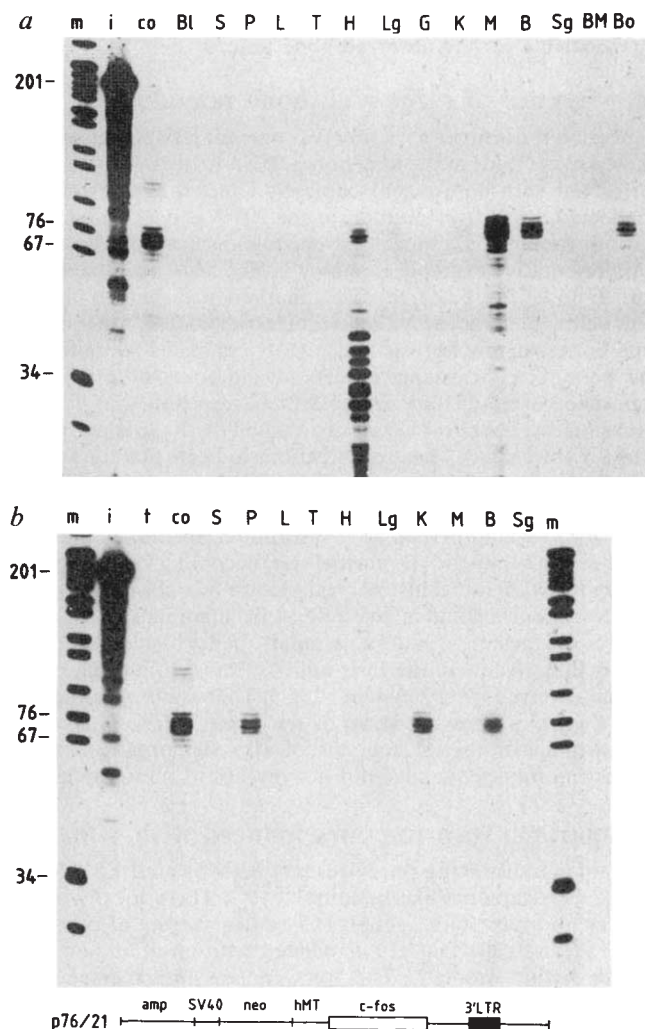


Fig. 2 MT-*c-fos* expression in various tissues of a F_1 offspring from mouse 209-4 (*a*) and 211-5 (*b*) after $CdCl_2$ induction. RNA analyses and abbreviations as described in Fig. 1, except for Bl, blood; T, thymus; BM, bone marrow and Bo, bone in *a*. Bottom, injected construct p 76/21.

to develop to term. On average 16% of injected and transferred eggs developed and of those 23% contained variable numbers of exogenous DNA copies (determined by hybridization analysis of tail DNA; data not shown). A total of 10 founder animals that developed normally was obtained and 8 strains of transgenic mice were established carrying the exogenous *c-fos* sequences in the germ line.

Expression of MT-*c-fos*

Founder animals and F_1 offspring of eight independent transgenic strains were analysed for exogenous *c-fos* expression using a sensitive RNase protection assay, which detects a diagnostic 75-base fragment from the human metallothionein messenger RNA (Fig. 1*a*). When RNA from 10 organs of each mouse line was analysed, exogenous *c-fos* expression was undetectable, except for a low level of expression in the testis and pancreas of two mice. The absence of vector sequences in seven mouse lines carrying MT-*c-fos* did not increase the expression. Likewise, injection of mice with $CdCl_2$ had no effect on the expression of exogenous *c-fos* mRNA.

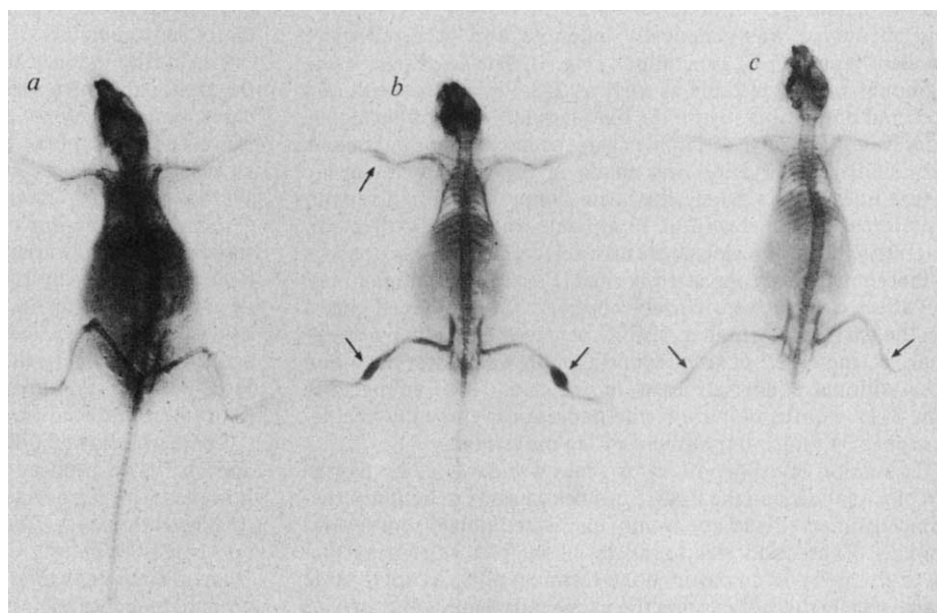
The absence of exogenous *c-fos* expression was further investigated by treatment with cycloheximide, a protein synthesis inhibitor, known to 'superinduce' *c-fos* expression *in vitro*, presumably by preventing *c-fos* mRNA degradation^{12,13}. When F_1 offspring were tested for expression after a 6-hour cycloheximide injection, significantly increased levels of exogenous *c-fos* transcripts were found in most tissues analysed, with the highest levels observed in testis and pancreas (Fig. 1*b*). This result suggests that the exogenous genes are transcribed in several tissues and that accumulation of *c-fos* mRNA—at least in some tissues—is prevented by post-transcriptional regulatory mechanisms.

A 3' modification allows stable *c-fos* expression

The mRNA-destabilizing sequence seems to be positioned in the 3' non-translated region of *c-fos*^{14,15}. To obtain transgenic mice with a high *c-fos* expression in many tissues we decided to use a construct in which this region was exchanged with a retroviral long terminal repeat (LTR) (the FBJ-MSV-LTR; see Fig. 2, construct p76/21)¹⁵. Five independent transgenic mouse families were established from eggs injected with p76/21. The founder animals contained 5–30 copies of exogenous DNA per cell (data not shown). Although three lines did not express p76/21 sequences (possibly because of the integration site or the presence of vector sequences), high levels of exogenous *c-fos*

Fig. 3 X-ray radiograph of normal and *c-fos*-expressing mice. *a*, Control C3H mouse; *b*, F_2 offspring 209-4-2-41 showing clearly visible swellings; *c*, F_2 offspring 209-4-2-44 without visible lesions. Arrows, bone abnormalities of different degrees. All three C3H mice were 15 weeks old.

Methods. Radiographs of anaesthetized mice were taken with a Siemens X-ray apparatus (B; 125/12/50R). The target-to-film distance was 105 cm and settings were 35 kV, 8 mA with 1.0 s exposure on Dupont, Cronex URF31 X-ray film.



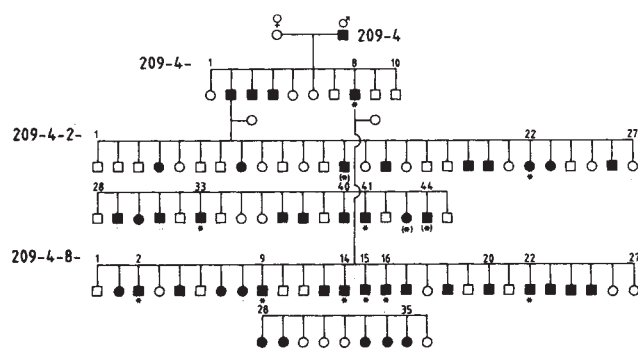


Fig. 4 A pedigree of MT-*c-fos* transgenic mice 209-4. Females are indicated by circles, males by squares and the transgenic animals by solid symbols. The founder male (209-4), 10 F_1 offspring and two F_2 sublines with 45 and 36 offspring are shown; * indicates visible bone abnormalities and (*) the mice with bone lesions detected only by X-ray or histological analysis.

RNA were detected in two families (209-4 and 211-5). The highest expression (determined by RNA blot analysis and RNase protection assay) was observed in pancreas, kidney, brain, heart and muscle. Lower, but readily detectable levels were found in lung and salivary gland in both strains tested (Fig. 2). This pattern of expression of the MT-*c-fos* gene differs from mouse MT-fusion genes, which are often highly expressed in liver²⁵. Upon induction with $CdCl_2$, the steady-state RNA level increased nearly 10-fold in most tissues, except for the liver in mice of both strains. These data show that the 3' deletion in the *c-fos* gene and the addition of a retroviral LTR is both necessary and sufficient for stable expression in a variety of tissues *in vivo* and that the MT-promoter-driven *fos* gene has a novel pattern of expression.

Abnormal bone formation

Having established two independent mouse strains expressing exogenous *c-fos* in several tissues, we examined whether over-expression of *c-fos* had biological consequences. We first noted that one young animal (209-4-8) in the F_1 generation of 209-4 was slightly smaller than normal, with visible swellings on the tibiae of both hind legs. These bone lesions are clearly visible on the X-ray picture shown in Fig. 3b (compare with Fig. 3a, wild-type mouse). Breeding analysis of two independent sublines originating from males 209-4-2 and 209-4-8 revealed that this phenotype was genetically inherited and occurred quite frequently in both F_2 generations (Fig. 4). The swellings on the leg bones were detectable as early as 2–3 weeks after birth and occurred predominantly on the hind legs in ~40% of the males, whereas only ~20% of females showed such abnormalities. A very similar observation was made in the second transgenic mouse line (211-5), where the same symptoms arose also with a preference for males (4 of 13 animals were visibly affected). Interestingly, bone lesions were also detectable by X-ray analysis in those mice which appeared normal (Fig. 3c). In both families, the affected mice were closely observed for prolonged times, but no swelling, whether visible or revealed only by X-ray analysis, increased in size beyond the 4th week after birth and no additional abnormalities were detected. These animals are now 8–11 months old, are fertile and appear clinically healthy except for a minor impairment of leg movement.

Expression of exogenous *c-fos* genes was analysed by Northern blot analysis and the RNase protection assay in both macroscopically affected and apparently unaffected tibiae from several animals. Expression was found in all samples suggesting that the *c-fos* gene is active in bone-forming cells, as total bone marrow cells did not express the exogenous genes (Fig. 2a). As *c-fos* expression in bone is normally barely detectable¹⁷, we

conclude that the deregulated expression of *c-fos* in bone cells is responsible for the observed bone lesions.

Interference of *c-fos* with bone remodelling

Generally, the lesions were slowly progressing swellings usually located at the ends of the long bones. Histological analyses were performed with both macroscopically affected and apparently unaffected tibiae from animals of the 209-4 and 211-5 families. In both families, a similar pathohistological picture emerged. A representative example is shown in Fig. 5b, d (compare with normal bone; Fig. 5a, c), revealing marked disturbances of bone remodelling, characterized by osteoclast-mediated bone resorption, bone marrow fibrosis and greatly enhanced formation of new bone. The coexistence of lamellar old bone and osteoblast-deposited osteoid and irregular, woven bone is typical. Occasionally, chondroblasts and confined cartilage islands were found in the lesions. These observations indicate that the normal processes involved in bone morphogenesis occur in the bone lesions, but remodelling occurs in an unordered way, with much more bone formation than bone resorption. Eventually the bone tissue expands beyond the normal size, becoming visible lesions. A practically identical histological picture was observed in both apparently normal and affected tibiae from animals of the same family, suggesting that a similar perturbation of bone remodelling occurs in the long bones of most, if not all, of the progeny derived from 209-4 and 211-5. The visible swellings are thus a more progressed stage of the same lesion. It is worth noting that histological analyses of all other organs of *c-fos*-expressing transgenic mice did not reveal any abnormalities.

Comparison with tumours induced with *v-fos*

Both *v-fos*-transducing retroviruses, FBJ-MSV and FBR-MSV, induce osteosarcoma-like tumours^{9,10,26,28}. These locally invasive, but non-metastatic, neoplasms consist largely of primitive mesenchymal cells that are associated with an abundant connective tissue stroma²⁹. The tumours are interspersed with islands of alkaline phosphatase-positive osteoblastic cells and foci of bone and cartilage formation, considered as areas of evolutionary metaplasia^{29,30}. The *v-fos*-induced tumours usually arise from periosteal cells of sternum, ribs, vertebrae and long bones or from the peritoneum and inguinal muscles and have been described as chondro-osseous neoplasms²⁹ or pleomorphic fibroblastic sarcomas³⁰ of low-grade malignancy. The reason for the specificity of virus-induced tumour type is not clear but could, for instance, be due to restriction of infection to specific cells, to lack of expression in certain infected cell types or to a target cell specificity of the *fos* gene product.

The virally induced tumours and the bone lesion detected in the transgenic mice show some intriguing similarities. In both cases, bone represents the (or a) target tissue and the resulting lesions consist, at least in part, of bone precursor (osteoblastic) as well as bone and cartilage-forming cells with areas of newly synthesized bone. These observations suggest that both the bone-associated virally-induced tumours and the bone lesion in the transgenic animals arise from a multipotent bone precursor cell believed to be closely associated with bone vessels. As expression of exogenous *c-fos* in the transgenic mice occurs in a variety of tissues other than bone without any obvious biological consequences, it is very likely that the observed specificity of the virus-induced tumour-type is due to a target cell specificity of *fos* protein-induced transformation.

There are marked differences between virus-induced tumours and the *c-fos*-induced bone lesions in the transgenic mice, however. Most importantly, the latter lesions are not malignant; their growth ceases a few weeks after birth, they are not invasive and consist primarily of differentiated bone-synthesizing cells. Primitive mesenchymal cells, present in large numbers in the virus-induced sarcomas^{29,30}, are practically absent from the non-malignant lesions. Several explanations which are not mutually

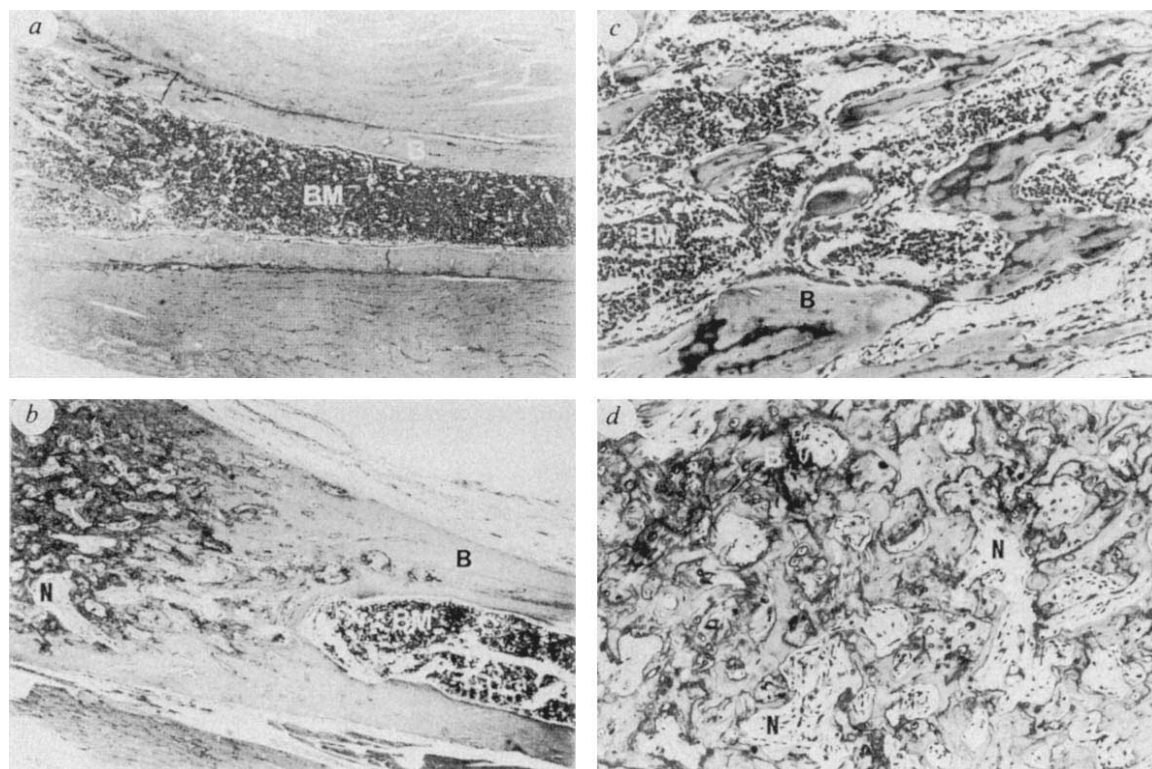


Fig. 5 Histology of bone lesions in the tibia of transgenic mouse 209-4-2-13. *a, c*, Different magnifications of a histological section showing normal bone development; *b*, low magnification of the affected tibia from 209-4-2-13. The changes are localized in the metaphysis, which contains trabecular bone and appears markedly enlarged. Between the bone trabeculae, bone marrow fibrosis and new bone formation is obvious. *d*, High magnification of the affected bone tissue. A characteristic mosaic pattern of bone development is clearly visible. The newly formed woven bone appears as slightly stained region between the bone lamella showing morphological features of bone resorption. B, bone; BM, bone marrow; N, new bone.

Methods. Standard 3- μ m sections of the tibia were made after fixation in 10% formaldehyde, decalcification and embedding in paraffin. The sections are stained with haematoxylin and eosin. Microscopic magnification $\times 30$ and $\times 120$ respectively.

exclusive are possible. (1) Although the normal *c-fos* gene product can induce morphological transformation in the rat fibroblast cell line 208F²¹, the structurally different *v-fos* protein of FBJ-MSV shows a higher transformation efficiency (T. Jenuwein and R.M., unpublished data). In addition, the FBR-MSV p75^{ag-fos} gene product has undergone structural alterations that have greatly enhanced its transforming capacity and have activated its immortalizing potential *in vitro*³¹. Thus, if the structural alterations in the *v-fos* genes have increased their transforming properties *in vitro*, it would be conceivable that similar differences may exist *in vivo*. (2) It is known for several oncogenes that the extent of transformation is dependent on the level of their expression^{32,33}. It is possible that expression of *v-fos* in the sarcoma cells is higher than the level of *c-fos* expression in the bone lesion in the transgenic mice. But as the level of exogenous *c-fos* expression in individual cells is not known, the relevance of this point is hard to judge. (3) For induction of osteosarcomas, newborn animals are infected with complexes of sarcoma and helper viruses. A role of the helper virus component in the process of sarcomagenesis can, therefore, not be completely ruled out. (4) The *v-fos*-induced sarcomas may have a less stringent requirement for factors supporting the growth of bone cells. Thus, the *c-fos*-induced lesion may be able to expand as long as such factors are supplied, that is, during the phase of normal bone growth. Such factors may act on the balance between bone formation and bone resorption. Interestingly, the *c-fos*-induced lesions cease to expand at the time when the animal reaches its adult size. The lesions may thus be the result of the cooperative action of exogenous *c-fos* expression and growth factors engaged in the regulation of this balance. This could also explain the observed specificity of the *c-fos*-

induced lesion for the long bones of the legs, as these bones normally show the highest degree of cell proliferation³⁴ and may thus be supplied with higher concentrations of growth factors.

Additional investigations are in progress to answer some of the open questions raised above and study the function of *c-fos* in development. We have recently produced transgenic mice carrying the *c-fos* gene linked to the H-2K^b promoter, which should direct expression in many different cell types, including the haematopoietic lineages, where we did not obtain expression with the MT-*c-fos* construct. In one transgenic mouse line carrying the H-2-*c-fos* gene with the 3' LTR modification on a different genetic background, multiple swellings were visible on the bones of all four legs. In contrast, several mice carrying H-2-*c-fos* genes without the 3' modification developed no obvious bone lesions (data not shown). These data confirm our data obtained with MT-*c-fos* and indicate that the perturbation of bone development is not specific for the promoter region and the mouse strain used, but requires the modification in the 3' non-coding region of the gene. Several retroviral vectors carrying *v-fos* genes are also being introduced into early embryos to study the mechanism of *fes*-induced sarcomagenesis.

Transgenic mice carrying other oncogenes

From a number of studies using transgenic mice and oncogenes, it is clear that most oncogenes predispose the transgenic animals to tumour formation, but secondary events seem to be required for the onset of oncogenesis⁷. In the case of targeted expression of SV40 T antigen to the β cells in the islets of Langerhans, all islets showed high level expression, yet only a few were transfor-

med to give rise to solid, encapsulated, vascularized tumours³⁵. Similarly, only one of the ten mammary glands in mice carrying a MMTV-*myc* gene developed mammary tumours, and only after the second (or later) pregnancy³⁶. In a recent study, again using the MMTV-*myc* fusion gene, expression was found in a wide variety of tissues leading to an increased incidence of different types of tumours, but also requiring secondary events to induce malignancy³⁷. The B-cell lymphomas produced by immunoglobulin-enhancer-*c-myc* fusion genes are generally clonal, implying a requirement for secondary event(s) to induce tumorigenesis⁸. Finally, SV40 and bovine papilloma virus I cause brain and skin tumours, respectively, after long latency periods, suggesting the requirement of additional steps for the development of neoplasia^{38,39}.

In contrast to all these previous observations, the *c-fos*-induced lesions develop very early, (with the onset of bone

development) and in many cases become visible shortly after birth. It is therefore unlikely that secondary genetic changes are involved. The formation of these lesions rather seems to be a direct consequence of deregulated *c-fos* expression. It can also be largely excluded that the *c-fos*-induced bone changes represent preneoplastic lesions, as no progression was observed over a period of up to 10 months in several animals. This study therefore represents a successful attempt in mammals to interfere with a specific developmental process by perturbing the regulation of a proto-oncogene. After submission of the manuscript, Langdon *et al.* have recently shown that deregulated *c-myc* expression interferes with the proliferation and maturation of lymphoid cells in transgenic mice⁴⁰.

We thank Willy Schoorl for careful handling of our mouse colony, Anne Walter for typing the manuscript and Mrs Schmiddhammer for taking the X-ray pictures.

Received 2 October; accepted 20 November 1986.

1. Bishop, J. M. & Varmus, H. E. in *Molecular Biology of Tumor Viruses, RNA Tumor Viruses*, 2nd edn. (eds Weiss, R. A. *et al.*) 999–1108 (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1982).
2. Varmus, H. E. *A. Rev. Genet.* **18**, 553–612 (1984).
3. Müller, R. *Trends Biochem. Sci.* **11**, 129–132 (1986).
4. Weinberg, R. A. *Trends Biochem. Sci.* **9**, 131–133 (1984).
5. Marshall, C. in *Molecular Biology of Tumor Viruses, RNA Tumor Viruses II* (eds Weiss, R. *et al.*) 487–558 (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1985).
6. Wagner, E. F. & Stewart, C. L. in *Experimental Approaches to Mammalian Embryonic Development* (eds Rossant, J. & Pederson, R. D.) 509–549 (Cambridge University Press 1986).
7. Hanahan, D. in *Oncogenes and Growth Control* (eds Kahn, P. & Graf, T.) 349–363 (Springer, Heidelberg, 1986).
8. Adams, J. R. *et al.* *Nature* **318**, 533–538 (1985).
9. Müller, R. *Biochim. biophys. Acta* **823**, 207–225 (1986).
10. Verma, I. M. *Trends Genet.* **2**, 93–96 (1986).
11. Greenberg, M. E. & Ziff, E. M. *Nature* **311**, 433–438 (1984).
12. Kruijer, W., Cooper, J. A., Hunter, T. & Verma, I. M. *Nature* **312**, 711–716 (1984).
13. Müller, R., Bravo, R., Burckhardt, J. & Curran, T. *Nature* **312**, 716–720 (1984).
14. Müller, R. & Wagner, E. F. *Nature* **311**, 438–442 (1984).
15. Rütter, U., Wagner, E. F. & Müller, R. *EMBO J.* **4**, 1775–1781 (1985).
16. Müller, R., Verma, I. M. & Adamson, E. D. *EMBO J.* **2**, 679–684 (1983).
17. Müller, R., Müller, D. & Guilbert, L. *EMBO J.* **3**, 1887–1290 (1984).
18. Kreipe, H. *et al.* *Differentiation* (in the press).
19. Müller, R., Müller, D., Verrier, B., Bravo, R. & Herbst, H. *EMBO J.* **5**, 311–316 (1986).

20. Karin, M., *et al.* *Nature* **308**, 513–519 (1984).
21. Miller, A. D., Curran, T. & Verma, I. M. *Cell* **36**, 51–60 (1984).
22. Meijlink, F., Curran, T., Miller, D. A. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4987–4991 (1985).
23. Treisman, R. *Cell* **42**, 889–902 (1985).
24. Shaw, G. & Kamen, R. *Cell* **46**, 659–667 (1986).
25. Brinster, R. L. *et al.* *Cell* **27**, 223–231 (1981).
26. Finkel, M. P., Biskis, B. O. & Jinkins, P. B. *Science* **151**, 698–701 (1966).
27. Finkel, M. P. & Biskis, B. O. *Prog. exp. Tumor Res.* **10**, 72–111 (1968).
28. Finkel, M. P., Reilly, C. A., Biskis, B. O. & Greco, I. L. *Proc. Symp. Colston Res. Soc.* **24**, 353–366 (Butterworths, London, 1973).
29. Ward, J. M. & Young, D. M. *Cancer Res.* **36**, 3985–3992 (1979).
30. Price, C. H. G., Moore, M. & Jones, D. B. *Br. J. Cancer* **26**, 15–27 (1972).
31. Jenuwein, T., Müller, D., Curran, T. & Müller, R. *Cell* **41**, 629–637 (1985).
32. Parker, R. C., Varmus, H. E. & Bishop, J. M. *Cell* **37**, 131–139 (1984).
33. Spandidos, D. A. & Wilkie, N. M. *Nature* **310**, 469–475 (1984).
34. Sissons, H. A. in *The Biochemistry and Physiology of Bone*, 2nd edn (ed. Bourne, G. H.) 145–175 (Academic, New York, 1972).
35. Hanahan, D. *Nature* **315**, 115–122 (1985).
36. Stewart, T. A., Pattengale, P. K. & Leder, P. *Cell* **38**, 627–637 (1984).
37. Leder, A., Pattengale, P. K., Kuo, A., Stewart, T. A. & Leder, P. *Cell* **45**, 485–495 (1986).
38. Brinster, R. L. *et al.* *Cell* **37**, 367–379 (1984).
39. Lacey, M., Alpert, S. & Hanahan, D. *Nature* **322**, 609–612 (1986).
40. Langdon, W. Y., Harris, A. W., Cory, S. & Adams, J. M. *Cell* **47**, 11–18 (1986).
41. Chirgwin, D. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. *Biochemistry* **18**, 5294–5299 (1979).
42. Melton, D. A. *et al.* *Nucleic Acids Res.* **12**, 7035–7056 (1984).

LETTERS TO NATURE

A determination of the white-dwarf masses in wide binary radio-pulsar systems

G. J. Savonije

Astronomical Institute Anton Pannekoek, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

The recent discovery of two new binary pulsars, PSR1831–00 (ref. 1) and PSR1855+09 (ref. 2), brings the total number of these objects to seven. By observations of pulse arrival time one can determine the mass function and $a_1 \sin i$, the projected orbital separation of the pulsar from the system's centre of gravity. By adopting a mass for the pulsar the mass of the secondary can be calculated as a function of the unknown orbital inclination angle i . Here I show that one can make use of the fact that four of these binary systems appear to fit neatly into the evolutionary scheme for bright X-ray sources described by Webbink *et al.*³. This implies that one can apply another constraint to the orbital solution, namely that the (sub)giant progenitor of the present white dwarf secondary should have filled its Roche lobe while spinning up the pulsar companion by mass transfer. With this extra constraint one can determine the mass of the present white dwarf in these systems^{4,5,6}. It also allows one to set an upper limit of $\sim 1.2 M_\odot$ to the mass of the neutron star PSR1855+09.

The currently known seven binary pulsars can be divided into two groups⁷, the 'PSR1913+16 class' consisting of systems with

relatively short orbital periods, high mass functions and very eccentric orbits, and the 'PSR1953+29 class' consisting of systems with long orbital periods, low mass functions and circular orbits (see Table 1). The systems in the first class are a direct result of 'spiral-in' evolution, and the systems in the second class are the result of subsequent 'spiral-out' evolution caused by mass transfer from a (sub)giant to a more massive compact companion. The tidal mass transfer continues until the envelope of the (sub)giant is depleted (by the mass loss as well as by the nuclear burning in the hydrogen shell source) and the secondary settles down as a helium white dwarf. The observed binary pulsar systems of the second class are believed to be in this terminal state.

I shall concentrate on the systems in this second class because their evolution is relatively well understood and can be described semi-analytically^{3–6,8,9}. It appears that the radius and luminosity of the (sub)giant are determined by the mass of the degenerate helium core, M_c , and are virtually independent of the mass in the envelope. The luminosity and radius are given to good approximation by the following polynomial fit³:

$$\ln(L_g/L_\odot) = a_0 + a_1 y + a_2 y^2 + a_3 y^3 \quad (1)$$

$$\ln(R_g/R_\odot) = c_0 + c_1 y + c_2 y^2 + c_3 y^3 \quad (2)$$

for $-0.4 \leq y \leq 0.6$, where $y = \ln(M_c/0.25 M_\odot)$ and M_c is the mass of the degenerate helium-core of the (sub)giant. As the systems under study were born near the galactic plane¹, we adopt the values for a_i and c_i that correspond to a normal solar metallicity of $Z = 0.02$.

Table 1 Currently observed binary radio pulsars¹ divided into two classes after van den Heuvel⁷

PSR	P (ms)	$\log_{10} B$ (G)	$\frac{a_1 \sin i}{R_\odot}$	P_b (days)	e	$\frac{F(M_2, M_x)}{M_\odot}$
0655+64	195.6	10.0	1.8	1.03	<0.00005	0.0712
1913+16	59.0	10.3	1.0	0.32	0.6171	0.1322
2303+46	1,066.4	11.8	14.1	12.34	0.6584	0.2463
1855+09	5.4	8.5	4.0	12.33	0.00002	0.0052
1953+29	6.1	8.6	13.6	117.35	0.0003	0.0027
1831-00	520.9	<10.9	0.3	1.81	<0.005	0.00012
0820+02	864.9	11.5	70.0	1,232.4	0.0119	0.0030

The tidal mass transfer from the (sub)giant can be calculated from³

$$\dot{M} = A(c_1 + 2c_2y + 3c_3y^2)/(\frac{5}{3} - 2M_2/M_p)(M_2/M_c)L_g/L_\odot \quad (3)$$

where the constant $A = 1.45 \times 10^{-11} M_\odot \text{ yr}^{-1}$ follows from the energy production rate of shell hydrogen burning³ and L_g is given by (1). Equation (3) is valid until the envelope mass is depleted to a few per cent of the core mass, after which the extended envelope shrinks rapidly (on a thermal timescale) and the secondary settles down as a white dwarf and the mass transfer terminates. At this stage the binary is fairly wide with an orbital period as observed today.

This evolutionary model is supported by two facts:

(i) the observed small eccentricity of the orbit while one of the stars has gone through a presumably violent phase in which it became a neutron star. The small eccentricity can be explained if the secondary was much bigger in the past (a subgiant) and caused effective tidal interaction in the binary¹⁰.

(ii) the rapid rotation of the pulsar, which is generally assumed to be caused by accretion-induced spin-up. The (sub)giant progenitor of the white dwarf is ideally suited for this as it yields a high ($\geq 10^{-9} M_\odot \text{ yr}^{-1}$) and stable mass transfer rate over a long period. This could force a white dwarf progenitor of the pulsar to collapse and become a neutron star, which subsequently was spun up to millisecond periods^{3,6,8,9}.

Adopting this plausible evolutionary history for the class 2 binary pulsars yields the extra constraint, mentioned above, on the orbital solution, namely that the (sub)giant should have filled its Roche lobe¹¹ just before it became a white dwarf:

$$R_g = 0.462a_1(M_2/M_t)^{-2/3} \quad (4)$$

where B_g is given by (2) in which we have to substitute M_2 , the present mass of the secondary, for the core mass M_c , M_t is the total system mass and a_1 is the pulsar's orbital separation from the system's centre of gravity. If we now combine (4) with the known mass function of the binary system

$$(M_2 \sin i)^3 / M_t^2 = F \quad (5)$$

we deduce that the Roche lobe filling (sub)giant must have obeyed the following simple mass-radius relation just before its envelope was depleted:

$$R_g(M_2) = B(M_2/F)^{1/3} \quad (6)$$

where the constant $B = 0.462(a_1 \sin i)$ and the mass function F are both observable quantities. Because R_g is fully determined by $M_2 = M_c$ through (2), one can solve (6) for M_2 without knowing the orbital inclination or pulsar mass.

Figure 1 shows (6) graphically for the four class 2 binary pulsars studied here. It appears that the binary pulsars PSR0820+02 and PSR1831-00 contain white dwarf secondaries at the two extreme ends (with $M_2 \approx 0.16 M_\odot$ and $0.45 M_\odot$, respectively) of our polynomial radius fit (equation 2), represented in Fig. 1 by the dotted line. It is difficult to determine theoretical 'error bars' for the core-mass-radius relation as this depends on the variety of input physics used in stellar-evolution calculations. Fortunately, the relation is very steep,

so that a small change in core mass yields a large variation in radius.

By substituting the obtained value for M_2 into (1) and (3) one can calculate the (sub)giant's luminosity and mass transfer rate just before envelope depletion. To calculate $\dot{M}$ one has to adopt a value M_p for the pulsar mass. But it appears that the value of $\dot{M}$ is insensitive to the precise value of M_p as long as $M_2/M_p \ll \frac{5}{6}$. Table 2 shows the results of such calculations. The second column gives the calculated mass M_2 of the secondary stars, which may be the most accurately determined white dwarf masses at present. In an independent study Joss *et al.* obtained very similar values for the white-dwarf masses in the binary radio pulsars PSR0820+02, PSR1953+29 and PSR1855+09.

The mass transfer is expected to spin up the neutron star to roughly its equilibrium spin period¹²:

$$P_{eq} = 6 \text{ ms } (B_0)^{6/7} (R_6)^{8/7} (M_p)^{-5/7} (M_{17})^{-3/7} \quad (7)$$

where B_0 is the pulsar's surface magnetic field strength in units of 10^9 G, R_6 its radius in units of 10^6 cm, M_p its mass in $M_\odot$ and M_{17} the accretion rate in units of 10^{17} g s^{-1} .

After termination of mass transfer in the binary the neutron star becomes a radio pulsar that spins down as it radiates away its rotational energy. By setting the present spin period of the pulsar equal to the equilibrium period that corresponds to the calculated mass-transfer rate at the end of the mass-transfer phase (as listed in table 2), one can derive an upper limit to the neutron star's magnetic field strength at the beginning of the radio phase. I adopted the values $R_6 = 1$ and $M_p = 1.4$ in (7). For PSR0820+020 it was assumed that the accretion rate equalled the Eddington limit of $1.5 \times 10^{-8} M_\odot \text{ yr}^{-1}$. The calculated (crude) upper limits are shown in Table 2. By comparing with the currently 'observed' (from spin-down measurements) field strength in Table 1, the derived upper limits appear consistent.

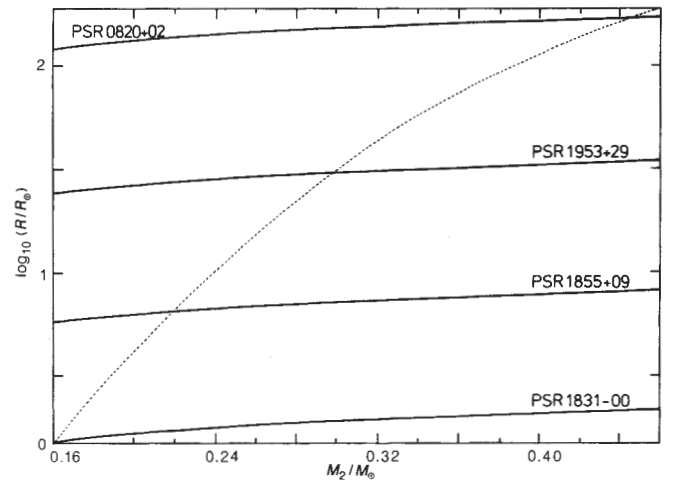


Fig. 1 A graphical determination of the white dwarf mass in wide binary radio-pulsar systems according to (6). The dotted line represents the polynomial fit (equation 2) to the core mass (M_2)-radius relation, and the continuous lines represent the term $B(M_2/F)^{1/3}$ for the four binary pulsar systems studied here.

Table 2 Calculated solution for the four low-mass binary radio pulsars

PSR	$M_2/M_\odot$	$R_g/R_\odot$	$L_g/L_\odot$	$\dot{M} (M_\odot \text{ yr}^{-1})$	$\log_{10} B_i (\text{G})$
1855+09	0.22	6.43	11.3	7×10^{-10}	≤ 9.0
1953+29	0.30	30.1	131.4	8×10^{-9}	≤ 9.6
1831-00	0.16	1.53	1.0	4×10^{-11}	≤ 10.7
0820+02	0.45	172	1,987	9×10^{-8}	≤ 12.3

The second column shows the mass of the white-dwarf secondaries. R_g , L_g and $\dot{M}$ are the radius, (intrinsic) luminosity and tidal mass-loss rate of the (sub)giant just before it became a white dwarf of mass M_2 . The last column gives an estimated upper limit for the neutron star's surface magnetic field strength at the beginning of its radio-pulsar phase.

The large differences between the derived field strengths for the various pulsars are probably due to the fact that the pulsars were born (by accretion-induced collapse) at different phases of the mass-transfer stage, so that the magnetic fields have

decayed to different degrees. That would imply that PSR1855+09 was formed relatively early, whereas PSR0820+02 was formed not long before the end of the mass-transfer stage. This seems consistent with the deduced short timescale for mass transfer and the observed relatively high orbital eccentricity ($e=0.0119$, ref. 1) in this system. It has been noted that the white dwarf in this binary is still relatively hot¹³, so that the mass-transfer phase terminated less than the white dwarf's cooling time ago. Although our analysis is independent of pulsar mass or inclination angle it can set limits to these values. From the mass function F we deduce that the pulsar mass M_p is given by $M_p = -M_2 + (F^{-1} M_2^3 \sin^3 i)^{1/2}$. Requiring that $\sin i \leq 1$ yields $M_p \leq -M_2 + (F^{-1} M_2^3)^{1/2}$. For three of the systems studied this yields a not very stringent upper limit $M_u > 2.8 M_\odot$ for the radio pulsar. For PSR1855+09, however, we obtain a much more interesting upper limit of only $\sim 1.2 M_\odot$. Unless this pulsar is a remarkably low-mass star, the solar system must be close to the orbital plane of this binary system.

Received 15 September; accepted 17 November 1986.

1. Dewey, R. J., Maguire, C. M., Rawley, L. A., Stokes, G. H. & Taylor, J. H. *Nature* **322**, 712-714 (1986).
2. Segelstein, D. J., Rawley, L. A., Stinebring, D. R., Fruchter, A. S. & Taylor, J. H. *Nature* **322**, 714-717 (1986).
3. Webbink, R. F., Rappaport, S. & Savonije, G. J. *Astrophys. J.* **270**, 678-693 (1983).
4. Savonije, G. J. *Nature* **304**, 422-423 (1983).
5. Paczynski, B. *Nature* **304**, 421-422 (1983).
6. Joss, P. C., Rappaport, S. A. & Lewis, W. *Astrophys. J.* (in the press).

7. van den Heuvel, E. P. J. in *Proc. IAU Symp* 125, Nanking 1986 (eds Helfand, D. J. & Huang, J. H.) (Reidel, Dordrecht, in the press).
8. Joss, P. C. & Rappaport, S. A. *Nature* **304**, 419-421 (1983).
9. de Kool, M., van den Heuvel, E. P. J. & Rappaport, S. A. *Astr. Astrophys.* **164**, 73-76 (1986).
10. Savonije, G. J. & Papaloizou, J. C. B. in *Interacting Binaries* (eds Eggleton, P. P. & Pringle, J. E.) 83-102 (Reidel, Dordrecht, 1985).
11. Paczynski, B. A. *Rev. Astr. Astrophys.* **9**, 183-208 (1971).
12. Davidson, K. & Ostriker, J. P. *Astrophys. J.* **179**, 585-598 (1973).
13. Kulkarni, S. R. *Astrophys. J.* **306**, L85-L89 (1986).

The formation of a magnetic-field-free cavity at comet Halley

W.-H. Ip & W. I. Axford

Max-Planck-Institut für Aeronomie, D-3411 Katlenburg-Lindau, FRG

One of the most important results of the Giotto mission to comet Halley was the discovery of a magnetic cavity surrounding the nucleus in which the magnetic field strength was apparently zero¹. The boundary of the cavity, the ionopause, occurred at 4,760 km (3,840 km) and the maximum magnetic field strength was ~ 55 nT (65 nT), thousands of km outside the rather sharply defined transition to zero field on the inbound (outbound) part of the trajectory, respectively. We show here with the aid of a simple model that such a magnetic field structure can be explained as being largely the result of ion-neutral friction² rather than simple pressure-balance alone^{3,4}.

Consider first the extreme situation in which the pressure of the plasma in the cometary ionosphere is negligible, the neutral cometary wind is directed radially away from the nucleus with speed W and the magnetic field direction is transverse to the flow. Assuming that only gradients in the radial direction are important and that there is no radial component of plasma flow in the region containing the magnetic field, the electric current is given by

$$j = en(R_i + R_e)W \quad (1)$$

where e is the charge on an electron, n the plasma number density and $R_i = m_i v_{in}/2eB$, $R_e = m_e v_{en}/eB^2$, B being the magnetic field strength, m_i and m_e the mass of an ion and an electron and v_{in} and v_{en} the ion and electron collision frequencies with neutrals, respectively. In addition to the transverse ion and electron motions which give rise to j , there is an electric field with a radial component $(R_e - R_i)WB$ and a transverse component $R_i R_e WB$, both perpendicular to the magnetic field vector.

Assuming a plane of symmetry such that

$$j = (1/4\pi) (1/r) \partial(rB)/\partial r \quad (2)$$

where r is the radial distance from the nucleus, it is a simple

matter to solve (1) and (2) for $B(r)$ for given values of the various parameters involved, notably the ion and neutral number densities, $n(r)$ and $N(r)$, respectively. We take $N(r) = Q/4\pi W r^2$, where Q molecules s^{-1} is the total sublimation rate from the nucleus.

Consider first the case in which the plasma density is determined by the condition of photochemical equilibrium so that $n(r) = \sqrt{(\beta/\alpha N(r))}$, where β is the photoionization rate of cometary H_2O molecules and α the dissociative recombination coefficient for the resulting ions and electrons. Defining

$$(B_m r_m)^2 = 4\pi(\beta/\alpha)^{1/2} (Q/4\pi W)^{3/2} (m_i K_i/2 + m_e K_e) W \quad (3)$$

with $v_{in} = K_i N$, $v_{en} = K_e N$, we find from (1) and (2) that

$$B = B_m(r_m/r)\sqrt{(1+2 \ln r/r_m)} \quad (4)$$

The constants have been chosen so that $B = B_m$, the maximum value of B , at $r = r_m$. The field strength is zero at a finite distance from the nucleus, namely $r_0 = r_m \exp(-0.5) = 0.61 r_m$. It is noteworthy that r_m is sensitively dependent on Q , the other quantities in (3) varying only weakly with distance from the Sun. If we take $\beta = 10^{-6} s^{-1}$, $\alpha = 10^{-7} cm^3 s^{-1}$, $W = 10^5 cm s^{-1}$, $Q = 8 \times 10^{29}$ molecules s^{-1} , $m_i K_i = 10^{-31} g cm^3 s^{-1}$ and neglect $m_e K_e$, we obtain $r_m B_m = 3.2 \times 10^{10}$ nT cm. Thus on the sunward side of the nucleus, where the magnetic field strength may be as large as 90 nT⁶, the maximum field strength should occur at a distance of about 3,500 km and the ionopause at about 2,200 km. To the extent that our analysis is valid on the flanks of the ionopause where it was encountered by Giotto, the corresponding distances (with $B_m = 60$ nT) are 5,300 km and 3,300 km.

As an alternative we also consider the case in which the plasma density is uniform, at least in the region where the current is strong; thus $n(r) = n_i$, with n_i a constant. Defining

$$B_m^2 r_m = n_i (m_i K_i/2 + m_e K_e) Q \quad (5)$$

we obtain

$$B = B_m \sqrt{(2r_m/r - r_m^2/r^2)} \quad (6)$$

and the magnetic field strength is zero at $r_0 = r_m/2$. With the above values of Q and $m_i K_i$ and with $n_i = 8 \times 10^3$, we obtain $r_m B_m^2 = 3.2 \times 10^{12}$ nT² cm. This implies that the ionopause should

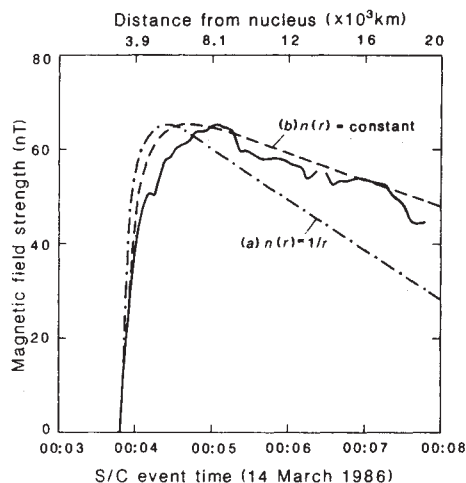


Fig. 1 A comparison of model magnetic field profiles with that observed by the Giotto magnetometer during the outbound pass of comet Halley¹. The case of constant $n(r)$ (a, plain broken line) provides the better fit to the observed profile than when $n(r) \propto 1/r$ (b, dotted broken line), but this could be partly a consequence of the fact that the pass was transverse to rather than along the comet-Sun direction. To fit the location of the magnetic field cut-off ($r_0 \approx 3,800$ km) and the peak field value ($B_m \approx 65$ nT), it is required that $Q = 1.1 \times 10^{30}$ molecules s^{-1} in case (a) (photochemical equilibrium) and $n_i = 6,000$ cm^{-3} with the same value of Q in case (b).

Note the S/C event time is in UT.

occur at a distance of about 2,000 km if $B_m = 90$ nT and at about 4,500 km if $B_m = 60$ nT.

The values of the various parameters we have used are not precise and any coincidence with the observed magnetic field configuration in the vicinity of the ionopause is to some extent accidental. However, it is possible to match the observations rather well, as Fig. 1 indicates, where the parameters have been adjusted so that the correct values of r_0 and B_m are achieved. The assumption that $n(r)$ is constant provides the better fit and is more consistent with the observed ion density distribution^{7,8}.

We have neglected the plasma pressure in the above, but this can be included provided that the temperature is known, or can be determined separately. Thus (1) should be modified to

$$j = en(R_e + R_i)W - (1/B) \partial P / \partial r \quad (7)$$

where $P = nk(T_i + T_e)$ is the pressure, k is Boltzmann's constant and T_i and T_e are the ion and electron temperatures, respectively. For a given $P = P(r)$, we can integrate (7) and obtain a modified form for the magnetic field. In particular, if $P(r) = P_1$ for $B = 0$ and $P(r) = P_2$ otherwise, with $P_1 > P_2$, then the solution proceeds as before but the ionopause occurs where $B^2/8\pi = P_1 - P_2$ as would be expected. In the limit, where the effects of friction are negligible, the solution is easily found to be $B = B_m(r_m/r) H(r - r_m)$, with $H(r)$ the step function and $B_m^2/8\pi = P_1 - P_2$. One can allow for more general forms for $P(r)$ but this is of little value unless an appropriate model is used, in particular for T_e .

We suggest that the ionopause observed surrounding the nucleus of comet Halley must have been principally the result of friction² rather than a difference in plasma pressure^{3,4}. This is supported by the good qualitative and even quantitative agreement outlined above and also the apparent macroscopic stability of the observed structure and the difficulty of accounting for the necessary pressure difference required otherwise. An ionopause determined by pressure balance alone is highly susceptible to Kelvin-Helmholtz and Rayleigh-Taylor instabilities, which should lead to an irregular and diffuse boundary with internal 'flux ropes' as seen at Venus, for example¹⁰. The stability analyses made so far do not allow for photochemical and compressibility effects, which could be of some significance, but

most importantly they do not allow for the effects of friction self-consistently in the configuration before it is disturbed^{11,12}. We have not yet succeeded in finding any instability of the configuration described by (4) or (6) and note that even on the microscopic level little is to be expected as the differences in the electron and ion drift motions contributing to j are comparable with the ion acoustic wave speed, for example, only in a region about 1 km thick where the field is weak and the current is a maximum.

To make a significant contribution to the overall stress balance the plasma pressure and especially the pressure differences must be of order $B_m^2/8\pi$. With $B_m \approx 60$ nT this implies that the product $n(T_e + T_i)$ must be $\sim 10^8$ K cm^{-3} . With the above values for the parameters involved, the assumption of photochemical equilibrium suggests that the plasma number density at the ionopause was about 3,000 cm^{-3} , requiring a rather high electron temperature of about 30,000 K, assuming that $T_i \approx 300$ K as expected and as indicated by the observations⁷. In fact there is no direct information about the value of nT_e but it is reasonable to assume that the electron temperature within the ionopause is less than that in the region surrounding it because the density of neutrals, which are the main cause of electron cooling, is greater. Because there is also little change in the plasma density across the boundary of the diamagnetic cavity^{7,8}, it is difficult to see how plasma pressure differences could provide the desired effect. Based on existing models⁹ we would in fact expect the plasma pressure to be an order of magnitude less than the observed $B_m^2/8\pi$.

Note added in proof: After submission of the original manuscript we learnt that T. E. Craven, in an independent work (*Adv. Space Res.*, in the press), had obtained a solution in part similar to the present model.

Received 30 June; accepted 4 November 1986.

1. Neubauer, F. M. *et al. Nature* **321**, 352-355 (1986).
2. Ip, W.-H. & Axford, W. I. in *Comets* (ed. Wilkening, L. L.) 588-634 (University of Arizona Press, Tucson, 1982).
3. Schmidt, H. U. & Wegman, R. in *Comets* (ed. Wilkening, L. L.) 538-560 (University of Arizona Press, Tucson, 1982).
4. Wallis, M. K. & Dryer, M. *Astrophys. J.* **205**, 895-899 (1976).
5. Axford, W. I., Cunnold, D. M. & Gleeson, L. J. *Planet. Space Sci.* **14**, 909-919 (1966).
6. Riedler, W., Schwingschuh, K., Yeroshenko, Ye. G., Styashkin, V. A. & Russell, C. T. *Nature* **321**, 288-299 (1986).
7. Balsiger, H. *et al. Nature* **321**, 330-334 (1986).
8. Krankowsky, D. *et al. Nature* **321**, 326-329 (1986).
9. Mendis, D. A., Houpsis, H. L. F. & Marconi, M. L. *Fundamentals cosm. Phys.* **10**, 1-380 (1985).
10. Elphic, R. C. *et al. J. geophys. Res.* **85**, 7679 (1980).
11. Galeev, A. A. & Lipatov, A. S. *Adv. Space Res.* **4**, 229-257 (1984).
12. Ershkovich, A. I. & Mendis, D. A. *Astrophys. J.* **302**, 849 (1986).

Atomic-scale surface modifications using a tunnelling microscope

R. S. Becker, J. A. Golovchenko & B. S. Swartzentruber

AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

The desire to modify materials on the smallest possible scale is motivated by goals ranging from high-density information storage to the purposeful transformation of genetic material. Here we report an atomic-scale modification of the surface of a nearly perfect germanium crystal, effected by the tungsten tip of a tunnelling microscope. We believe this to be the smallest spatially controlled, purposeful transformation yet impressed on matter and we argue that the limit set by the discreteness of atomic structure has now essentially been reached.

Significant progress in the field of surface physics over the past 25 years has provided the tools (ultra-high-vacuum (UHV) environments, argon-ion sputtering, low-energy electron diffraction, Auger electron spectroscopy) to prepare crystal surfaces in a chemically pure and structurally homogeneous state. This homogeneity is manifest as a nearly perfect, periodic, two-

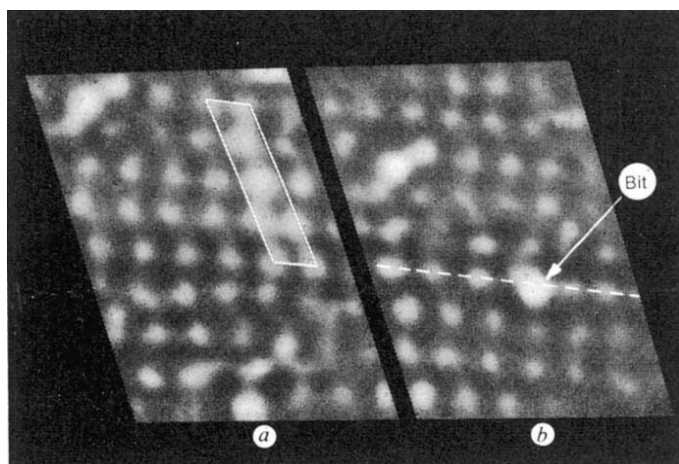


Fig. 1 Tunnelling images of a reconstructed (111) germanium surface covering a $50 \times 65 \text{ \AA}$ region. *a*, The surface before modification. A single unit cell of the $c-(2 \times 8)$ reconstruction is highlighted and a naturally occurring defect in the upper left hand corner serves as a registration mark. *b*, The same region after the surface modification. The displayed region is slightly translated due to thermal drift in the tunnelling microscope. The new bright spot near the centre of the picture is the impressed bit.

dimensional atomic array. The method of modifying the surface on an atomic scale is an extension of our recent studies of the electrical properties of semiconductor surfaces using a tunnelling microscope. A complete understanding of the procedures that have been effective requires further discussion of the instrument^{1,2}. The tunnelling microscope consists of a fine tungsten needle mounted on a set of piezoelectrically controlled manipulators. The ultimate resolution that may be obtained with the instrument ($\sim$ several angstroms) indicates that the needle's tip is terminated by a few or perhaps even just a single atom. The manipulators are capable of moving the needle in all three dimensions continuously with fine control of $0.1 \text{ \AA}$.

The distance between the tip and a sample surface is maintained by one of these manipulators which is controlled by a feedback network. This feedback network senses the small tunnelling current induced to flow across the atomic-scale gap between tip and sample by a small bias voltage, compares that value with a desired reference and uses the time-integrated error signal to drive the piezo manipulator to adjust the gap so that the tunnel and reference current agree. Knowledge of the piezo manipulator calibration constant, in angstroms per volt, together with the closed-loop integrated error signal applied to it, determines the actual tip position in a direction normal to the surface. When desired, lateral motion of the tip across the surface is simply adjusted with two orthogonal non-servo controlled piezoelectric transducers. A tunnelling image of a surface is a

display of the feedback-controlled tip height as a function of its lateral position.

Detailed studies of the electrical properties of the tip-sample tunnel junction can be informative in attempts to interpret the tunnelling images obtained with the microscope³. While extending these studies to the case of the germanium surface we discovered that certain bias conditions resulted in a stable transformation in the atomic structure of the surface just below the tip. The transformation, controlled, localized and on an atomic scale, is the transformation we referred to above.

Figure 1*a* shows a tunnelling image of UHV-prepared Ge(111) surface before transformation. The tip was biased at -1.0 V relative to the sample and a tunnelling current of 20 pA was used. The data are displayed as a grey scale map where each lateral data point is represented by a corresponding pixel and the grey level of each pixel has been mapped from the recorded tip height. Light regions are higher (the tip had to withdraw to maintain 20 pA of tunnel current) than dark ones. Figure 1 corresponds to a region $50 \times 65 \text{ \AA}$ in lateral extent most of which is covered by the $c-(2 \times 8)$ reconstruction⁴. A unit cell is indicated in Fig. 1*a*. Also seen are naturally occurring isolated defects where the periodicity of the surface lattice is broken. The imperfection in the upper left-hand corner will serve as a reference or registration marker below.

The essentially perfect central region of the Fig. 1 was chosen as the place to write an atomic-scale 'bit'. This was accomplished by raising the tip to surface bias to -4.0 V at a quiescent tunnel current of 20 pA while the tip remained laterally stationary over the site chosen for modification. Successful modification of the surface is indicated electrically by a rapid withdrawal of the tip from the surface by $\sim 1 \text{ \AA}$. The bias was then reduced to the value used for taking tunnelling topographs ($\sim 1 \text{ V}$) and an image, shown in Fig. 1*b*, of the entire region surrounding the impressed bit was obtained by the same procedure used to obtain Fig. 1*a*. The same naturally occurring disruption of the lattice periodicity in Fig. 1*a* may be easily identified and used to register the two images relative to one another. The main feature of interest is the small isolated new protrusion in a previously pristine region of the surface. This transformation, indicated by the arrow in Fig. 1*b*, is the smallest spatial rearrangement of surface atoms we have observed to date.

The scale of the transformation can be appreciated in Fig. 2 which shows before and after plots of the surface height along the line shown in Fig. 1*b* which passes directly through the bit position. The small modulations characteristic of the undisturbed surface crystallinity in Fig. 2*a* are augmented by a $1\text{-\AA}$ high bump with a full width at half maximum of $8 \text{ \AA}$ in Fig. 2*b*.

Figure 3 shows a plot of the typical electromechanical behaviour of the tip-sample tunnel junction with the feedback loop closed and a demanded tunnel current of 20 pA . The general character of the curves reflects the fact that, for increasing bias, the tip must be withdrawn to maintain constant tunnel current. Curve *ab* corresponds to a voltage scan where no transformation of the surface was induced. Curve *acde* shows

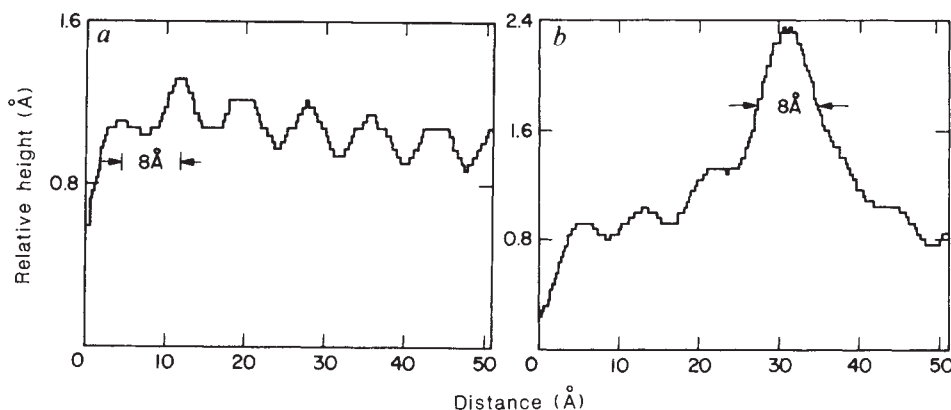


Fig. 2 Scan of the surface height: *a*, before; *b*, after impression of the bit. The scan is along the dotted line in Fig. 1*b* which passes through the bit position.

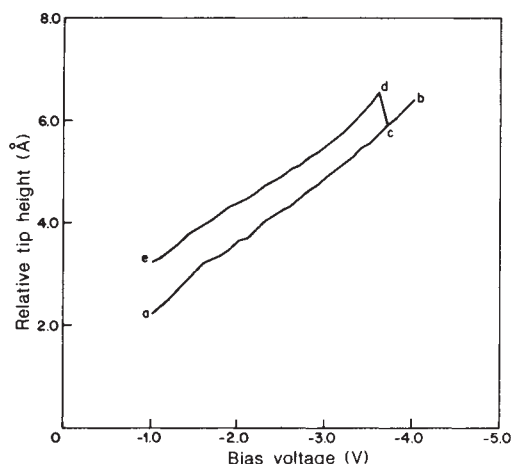


Fig. 3 Electromechanical junction response.

the situation under which an irreversible transformation was induced at *c* corresponding to a tip withdrawal of 1 Å *cd* which is maintained as the voltage is reduced *de*.

A few additional experimental observations are relevant to the discussion. First, we have never observed the transformation described above on silicon (111) surfaces even though tunnel currents were increased to several nanoamps and bias voltages as high as 20 V were used. Second, although this transformation could be repeated at different places on the surface not every attempt was successful. In some experiments, >90% of the attempts were successful and in others <10% were. The success rate improved when the tip was 'recharged' by taking it to a far-removed place on the surface and bringing it into contact with the surface.

Because the bit appears as added matter above the surface, one immediately questions its origin. One possibility is the tip itself. Germanium atoms previously transferred to the tungsten tip by contact with the sample may be redeposited under the electrical stimulation⁵ described above. Note that whereas stable silicides of tungsten exist, no germanium tungsten alloys are known. Germanium atoms may, therefore, be much less tightly bonded to the tungsten tip than silicon which would account for the success on germanium surfaces. It also seems reasonable to attribute the variability of results to variability in the state of the unobserved and poorly characterized tip rather than the uniform sample surface. The ability to recharge the tip also suggests it is the matter source. On the other hand, a reconfiguration of atoms of the equilibrium surface reconstruction may be stimulated by the high fields and current densities involved. We prefer the former explanation for the data and will discuss this in detail elsewhere. Nevertheless we believe the second mechanism may have an important role under appropriate circumstances.

We are uncertain about the actual atomic structure of the surface bit. In particular, the observed lateral dimension is certainly an upper limit to the actual size as the tunnelling images are a kind of convolution between the tip and surface geometry. To complicate matters the explanation for the unmodified *c*-(2×8) germanium surface is not completely understood so the details of the bit structure are difficult to anticipate. The understanding of nonequilibrium surface defect structures of the type we have induced represent a major new challenge to the materials and surface physics community.

We are convinced that a new rich field lies ahead in the study of atomic-sized structures on surfaces (see refs 6 and 7). The final rewards may be high-density memories, new devices whose electrical properties are dominated by quantum-size effects, or the modifications of a single self-replicating molecule on a surface. Such endeavours will require a much more elaborate understanding and control over materials and mechanisms than

we have demonstrated. Nevertheless, we hope to have shown that man can now manipulate a few chosen atoms for his own purposes.

Received 25 July; accepted 28 October 1986.

1. Binnig, G., Rohrer, H., Gerber, C. & Weibel, E. *Phys. Rev. Lett.* **50**, 120-123 (1983).
2. Golovchenko, J. A. *Science* **232**, 48-53 (1986).
3. Becker, R. S., Golovchenko, J. A., Hamann, D. R. & Swartzentruber, B. S. *Phys. Rev. Lett.* **55**, 2032-2034 (1985).
4. Becker, R. S., Golovchenko, J. A. & Swartzentruber, B. S. *Phys. Rev. Lett.* **55**, 2032-2034 (1985).
5. Hren, J. J. & Ranganathan, S. R. (eds) *Field Ion Microscopy* (Plenum, New York, 1968).
6. McCord, M. A. & Pease, R. F. W. *J. Vac. Sci. Technol.* **B4**, 86 (1986).
7. Ringger, M., Hidber, H. R., Schlögl, R., Oelhafen, P. & Guntherodt, H.-J. *Appl. Phys. Lett.* **46**, 832 (1985).

Inertial waves identified in the Earth's fluid outer core

K. D. Aldridge & L. I. Lumb

Department of Earth and Atmospheric Science, York University, Toronto M3J 1P3, Canada

Several inertial waves have been identified in the long period gravimetric data of Melchior and Ducarme¹. These hydrodynamical waves which only exist because of the Earth's rotation, must be in the Earth's fluid outer core. Close proximity of the observed periods to those predicted for a homogeneous fluid suggests the waves should be labelled inertial because buoyancy and compressibility effects must be small or self-cancelling. Some of the identified waves have azimuthal wavenumber one, consistent with their occurrence after large, deep earthquakes which probably excite the waves through a small perturbation in the Earth's rotation. Other waves are axially symmetric, consistent with a small change in the magnitude of the fluid core's rotation rate. All the waves decay more rapidly than would be expected for Ekman dissipation which suggests an additional dynamical damping associated with mantle-inner core coupling. The inertial waves identified here will serve as a precise tool for subsequent evaluation of models for core dynamics and the geodynamo.

The region of the Earth between the depths of about 2,887 and 5,142 km defines the Earth's outer core, which has long been recognized to be a fluid. The oscillatory response of this slightly compressible, electrically conducting, nearly homogeneous, contained, rotating fluid to small, long period perturbations will be dominated by the Coriolis force. Such rotationally determined behaviour of a contained fluid produces what are known as inertial waves². Previous investigations of inertial waves have centred on meteorological and oceanographic applications, although they had been recognized much earlier³. Experimental investigations have been more recent with excitation and detection in cylindrical, spherical and conical cavities⁴⁻¹¹. Theoretical work on inertial waves has attracted attention partly because the ill-posed nature of the boundary value problem brings the existence of continuous solutions into question¹².

Recently geophysical interest has centred on the dynamics of a realistic fluid outer core taking into account compressibility, stratification and self-gravitation. Despite the mathematical complexity considerable progress has been made¹³⁻¹⁶. Since it is our expectation that the behaviour of the fluid core at the periods observed by Melchior and Ducarme will be dominated by rotation, we have chosen to model the core accordingly. Thus by introducing what might be called the Poincaré Earth model, we expect to find some small differences between our predicted eigenfrequencies and those already observed for the core. In future work these differences will be used to determine properties of the core.

Table 1 Comparison of periods observed by Melchior and Ducarme with calculated inertial wave periods

Observed period (h)		Calculated period (h)	mode (n, m, k)
Mindanao	Hindu Kush	T_{nmk}	
15.3	15.2	15.832	(3, 1, 1)
13.9	14.0	14.003	(4, 1, 1)
13.2	13.2	13.247	(5, 1, 1)
—	15.6	15.626	(5, 1, 0)
14.4	14.3	14.403	(6, 1, 0)
13.6	13.6	13.720	(7, 1, 0)
13.4	13.4	13.290	(8, 1, 0)
14.6	14.6	14.551	(6, 1, 2)

Here, we compare the periods near peak power from the superconducting gravimeter with those of inertial waves of an incompressible, homogeneous, rotating fluid which completely fills a spheroidal cavity with a flattening, f , equal to that of the Earth's fluid core. Our interpretation of the observed periods associated with the Mindanao (20 November 1984) and Hindu Kush (30 December 1983) earthquakes are shown in columns one and two respectively of Table 1.

The inertial wave periods $T_{nmk} = 2\pi/\omega_{nmk}$ are related to the m roots^{17,18}

$$x_{nmk} = \frac{\omega_{nmk}/2\Omega}{[1 + \varepsilon(1 - (\omega_{nmk}/2\Omega)^2)]^{1/2}}$$

of

$$(1 - x^2) \frac{dP_n^k(x)}{dx} = k \left(\frac{1 + \varepsilon x^2}{1 + \varepsilon} \right)^{1/2} P_n^k(x),$$

where Ω is the Earth's rotation rate and $\varepsilon = 1/(1-f)^2 - 1$ and $f = 1/392$. For each n of the Legendre Polynomial P_n^k , the smallest value of m corresponds to the simplest radial structure and k is the azimuthal wavenumber. The calculated periods and modal numbers are given in columns three and four respectively of Table 1.

Agreement between the observed and calculated periods for inertial waves with $k = 1$, shown in the upper part of the table, is within about 3%. It is important to point out that the calculations given here do not include effects of the inner core. This apparently minor correction has proved to be difficult to obtain because the boundary value problem is ill-posed¹². Experimental measurements of the eigenfrequencies of inertial waves in spherical and spheroidal shells have been made^{4,5}, however, and they differ from those of the sphere by only a few percent for the spatially simple modes shown in Table 1. It appears, therefore, that the data of Melchior and Ducarme can be interpreted as the discovery of predominantly inertial waves in the Earth's fluid outer core.

Some additional properties of the observed spectra require further discussion. First, the spectral peaks show a small but measurable change in frequency with time while the waves decay. Wave periods depend only on geometry and rotation speed, so it is apparent that some small change in rotation speed must be occurring while the waves decay, because the geometry of the outer fluid core should remain the same on the time scale of the decay. Later we shall discuss the source of this change but it follows that any axisymmetric perturbation such as a change in the fluid's rotation speed, should excite axisymmetric waves⁶. A search of the data for axisymmetric waves with relatively simple spatial structure yielded the modes shown in the middle part of Table 1 along with their predicted values. Only the axisymmetric modes (3, 1, 0) and (4, 1, 0) have simpler spatial structure and their periods are outside the range reported in the observations.

A second noteworthy aspect of the observed waves is their relatively rapid decay. A plot of the log of the spectral amplitude over time gives an e-folding time for the decay of about 280 days. If the decay were simple Ekman dissipation the time scale of the decay would be

$$O(E^{-1/2}\Omega^{-1})$$

where $E = 7 \times 10^{-16}$, the Ekman number for the fluid core assuming simple Newtonian viscosity¹⁹ and $\Omega = 7.2921 \times 10^{-5} \text{ s}^{-1}$. This comes out to about 16,500 years which can only be reduced slightly by assuming a more realistic eddy viscosity. Although simple Ekman dissipation includes the efficient removal of energy through phase differences between the boundary layer and interior velocity fields, this process does not successfully model the observed rapid decays.

Some departure from the Poincaré model is required to account for the relatively rapid decays observed by Melchior and Ducarme. We note first that in the core the boundaries themselves may be moving and may also be deforming. In both these cases much more rapid decays would occur than would be found for stationary rigid boundaries but the eigenfrequencies themselves would be essentially unaffected. Another mechanism for the decay of the predominantly inertial waves is through magnetic coupling between the core and mantle. This type of coupling has been studied by Crossley¹³ and found to be ineffective at the periods discussed here.

The observed rapid decay of the inertial waves could be apparent rather than real. This would be the case if there were interaction among the waves while the decay took place. Indeed this coupling is found experimentally as verified by Stergiopoulos and Aldridge¹⁰. They found that the spin-over mode, (2, 1, 1) in the present notation, would gain energy from other modes during their free decay. Such an exchange of energy among modes might account for the $k = 2$ mode identified as (6, 1, 2) in the lower part of Table 1.

In the Earth's core interaction among modes would be further complicated by the possible movement of the mantle and inner core boundaries. In fact, displacement of the core-mantle boundary associated with the two earthquakes would be consistent with $k = 1$ and $k = 2$ azimuthal dependencies of the inertial waves.

One or more of the processes just given could account for the dissipation of the observed waves. To fix ideas for discussion of decay, excitation and detection of the waves, consider the following specific process. It is known that the mantle and the inner core are strongly coupled gravitationally²⁰ so that a small change in angular momentum of the mantle would be reflected by a larger fluctuation of the inner core's angular velocity. Such a perturbation of the inner core would excite inertial waves of azimuthal wavenumbers 0 and 1 if there were changes in both magnitude and direction of the inner core's velocity. Through conservation of angular momentum, changes in the magnitude and direction of the inner core's velocity would accompany corresponding changes in the angular momentum of the fluid outer core associated with each of the excited inertial waves. Differences in density between the inner core and fluid outer core are then responsible for the observed gravity anomalies.

The calculated inertial wave periods in Table 1 do not include the effect of an inner boundary. Experimental evidence^{4,5} for both the axisymmetric and non-axisymmetric inertial waves indicates a shift in wave periods of only a few percent for modes with simple spatial structure such as those in Table 1. Currently, calculations of inertial wave eigenfrequencies for a spherical shell are being performed in our laboratory. This complex mathematical problem cannot be solved in closed form but requires a variational method, for example, to obtain approximate eigenfrequencies²¹.

It is our expectation, based in part on our own experimental evidence, that calculated eigenfrequencies for a spheroidal shell will show only a small difference from those values given in

Table 1. We conclude that any residual discrepancy between observed wave periods and those predicted for a spheroidal shell will define models of a compressible, self-gravitating, stratified and possibly differentially rotating fluid core. In the case of weak stratification, for example, it is already known²² that the correction to inertial wave periods is $\sim(N/2\Omega)^4$ where N is the Brunt-Väisälä frequency, a measure of buoyancy effects.

It is hard to overestimate the importance of data like those obtained by Melchior and Ducarme in the study of the Earth's fluid core. More data are needed to refine estimates of periods²³. More stations are needed to resolve azimuthal wavenumbers. Greater use of very long baseline interferometry²⁴ will complement the gravimetric data in solving both of these problems. Finally, we note that the role of inertial waves in the geodynamo²⁵ will be elucidated as more waves are identified.

We thank Doug Smylie for several helpful and encouraging discussions. We have benefited from conversations with Tony Szeto and Barry Brennan. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Received 24 September 1986; accepted 3 December 1986.

1. Melchior, P. & Ducarme, B. *Phys. Earth planet. Interiors* **42**, 129–134 (1986).
2. Kelvin, Lord *Phil. Mag.* **10**, 155–168 (1880).
3. Bjerknes, V. & Solberg, H. *Ash. norske Vidensk. Akad.* **7**, 1–16 (1929).
4. Aldridge, K. D. *Mathematika* **19**, 163–168 (1972).
5. Aldridge, K. D. *Geophys. J. R. astr. Soc.* **42**, 337–345 (1975).
6. Aldridge, K. D. & Toomre, A. *J. Fluid Mech.* **37**, 307–323 (1969).
7. Beardsley, R. C. *Stud. appl. Math.* **49**, 187–196 (1970).
8. Fultz, D. *J. Met.* **16**, 199–208 (1959).
9. McEwan, A. D. *J. Fluid Mech.* **40**, 603–640 (1970).
10. Stergiopoulos, S. & Aldridge, K. D. *Phys. Earth planet. Interiors* **36**, 17–26 (1984).
11. Thompson, R. *J. Fluid Mech.* **40**, 737–751 (1970).
12. Stewartson, K. & Rickard, J. A. *J. Fluid Mech.* **35**, 759–773 (1969).
13. Crossley, D. J. *Phys. Earth planet. Interiors* **36**, 1–16 (1984).
14. Crossley, D. J. & Rochester, M. G. *Geophys. J. R. astr. Soc.* **60**, 129–161 (1980).
15. Crossley, D. J. & Smylie, D. E. *Geophys. J. R. astr. Soc.* **42**, 1011–1033 (1975).
16. Smylie, D. E., Szeto, A. M. K. & Rochester, M. G. *Rep. Prog. Phys.* **4**, 855–906 (1984).
17. Kudlick, M. D. thesis, M.I.T. (1966).
18. Greenspan, H. P. *The Theory of Rotating Fluids* (Cambridge University Press, 1969).
19. Gans, R. F. *J. geophys. Res.* **77**, 360–366 (1972).
20. Szeto, A. M. K. & Smylie, D. E. *Phil. Trans. R. Soc. A313*, 171–184 (1984).
21. Smylie, D. E. & Rochester, M. G. *Geophys. J. R. astr. Soc.* **86**, 553–561 (1986).
22. Friedlander, S. *Geophys. J. R. astr. Soc.* (submitted).
23. Rochester, M. G. & Crossley, D. J. *Geophys. Res. Lett.* (submitted).
24. Carter, W. E. & Robertson, D. S. *Scient. Am.* **255**, 46–54 (1986).
25. Olson, P. *J. geophys. Res.* **86**, 10875–10882 (1981).

Geometry, conditions and timing of off-axis hydrothermal metamorphism and ore-deposition in the Solea graben

Peter Schiffman*, Brian M. Smith†, Robert J. Varga† & Eldridge M. Moores*

* Department of Geology, University of California, Davis, California 95616, USA

† Unocal Science and Technology, PO Box 76, Brea, California 92621, USA

The recent discovery of fossil spreading axes along the northern margin of the Troodos ophiolite¹ provides a unique opportunity to compare a fossil seawater-hydrothermal system with systems of modern oceanic spreading centres^{2–6}. Whereas previous accounts of ophiolite hydrothermal systems^{7–13} including Troodos^{14–16} have elucidated many aspects of sea floor hydrothermal metamorphism, earlier workers have not placed ophiolite hydrothermal features within any specific spreading-centre structural framework. Most ophiolites exhibit complex histories of post-accretion deformation that mask or obliterate early structures related to seafloor spreading. The Troodos ophiolite is apparently still in the process of being 'emplaced'¹⁷ and most structural elements at Troodos appear to have originated beneath a Cretaceous sea floor¹. We have discovered an extensive area of intense seawater-hydrothermal altera-

tion within the Sheeted Dyke Complex of the Solea graben (Fig. 1). The metamorphic petrology and oxygen-isotope geochemistry of >100 dyke samples from this graben reveal the size and three-dimensional geometry of the hydrothermal system and yield estimates of temperature, salinity, oxygen-isotope composition and recharge-discharge geometry for fluids that transported the constituents of massive sulfide orebodies. Geometric discordance between the trends of the hydrothermally altered area and of earlier, seafloor spreading structures compels us to consider a post-spreading, off-axis origin for the orebodies of the Solea graben.

The Solea graben is the oldest and best-exposed of three asymmetric structural grabens identified along the northern margin of the Troodos ophiolite¹. Originally steep dykes and normal faults dip toward the graben axes, whereas originally horizontal lavas dip away. The eastern margin of the Solea graben is intruded by steep and/or northeastward-dipping dykes of the younger Ayios Epiphaniou graben (Fig. 1a). The progressive younging of grabens from west to east suggests discontinuous migration or 'jumping' of the Troodos spreading axis with time¹. Two of the largest massive sulphide deposits of the Troodos complex lie within the Solea tectonic graben (Fig. 1a, b): Skouriotissa (6 Mt of 2% Cu preproduction reserves) is located along the graben axis and Mavrovouni (15 Mt, 4% Cu) lies 3 km to the west¹⁸. The striking similarities between hydrothermal vent fragments in Troodos massive sulphides and fragments of present day 'black smokers' suggest that Troodos orebodies were formed by submarine hydrothermal systems similar to modern ones¹⁹. Studies of O-, H- and Sr-isotope systematics of hydrothermally altered rocks provide unequivocal evidence that Troodos orebodies formed when hot, upwelling fluids derived from Cretaceous seawater vented at the sea floor^{15,20}.

Our detailed field and analytical studies were undertaken to clarify relations between spreading structures, hydrothermal circulation and massive sulphide deposition within the Solea graben. Fig. 1 is a compilation of structural, petrographic, electron microprobe and whole-rock oxygen-isotope data from volcanic and dyke rocks from this graben. Fig. 1a depicts key hydrothermal mineralogical features of Solea rocks, including (1) the distributions of argillic (alkali feldspar+quartz+chlorite+/-pyrite) alteration and surrounding rocks containing secondary alkali feldspar, (2) the first appearance of epidote, (3) the complete albitization of primary plagioclase and (4) the distribution of granular epidotes (metasomatic epidote+quartz+/-chlorite rocks) that formed within hydrothermal zones with high fluid:rock ratio²¹.

Hydrothermal mineral zonation patterns within the Solea graben are strikingly different from those reported for many other ophiolites (for example, Del Puerto¹³, Horokanai²² and Sarmiento¹⁰), for the zonation patterns shown in Fig. 1a are largely non-conformable with the ophiolite pseudostratigraphy. For example, the nearly coincident epidote-in and primary plagioclase-out boundaries are near the base of the dyke complex along the graben axis, but continue upwards to a level high within the dyke complex along the western graben margin. The axis and narrow eastern margin of the graben contain subgreenschist facies dykes such that unalbitized plagioclase, unaltered clinopyroxene and randomly interlayered smectite-chlorite occur in samples close to the gabbro contact, which is interpreted to be a gently dipping detachment zone for planar and listric-normal, graben-forming faults¹. Along the eastern margin of the Solea graben, greenschist facies assemblages locally occur in dykes near the intrusive contact of the younger Ayios Epiphaniou graben (Fig. 1a).

The most extensively recrystallized dykes occur in the western Solea graben, where the Sheeted Dyke Complex displays two distinct styles of hydrothermal metamorphism: (1) nearly isochemical alteration to normal variance greenschist facies assemblages (albite+epidote+chlorite+/-actinolite) with retention of obvious diabasic igneous textures and (2) superim-

Table 1 Oxygen isotope and selected trace element data for dykes and lavas from the Solea graben, Troodos ophiolite

Lithology (no. of samples analysed)	$\delta^{18}\text{O}(\text{‰})$ versus SMOW)	Concentration (p.p.m.)*					
		Rb	Sr	Y	Zr	Cu	Zn
Epidosite dykes (6)	2.8–5.0	1 (0)	202 (165–275)	28 (18–35)	5 (0–31)	16 (0–49)	14 (0–37)
Greenschist dykes (5)	5.1–12.3	1 (0–6)	112 (72–214)	24 (13–49)	36 (0–84)	86 (12–277)	64 (0–287)
Subgreenschist dykes (7)	7.3–13.6	4 (0–21)	120 (49–128)	32 (23–45)	59 (0–107)	107 (48–261)	103 (60–181)
Lava flows (8)	8.8–14.4	23 (0–82)	115 (83–137)	26 (11–36)	54 (0–80)	73 (18–161)	93 (46–263)

* Mean (range below in parentheses).

posed, high-variance epidosite alteration in dykes pervasively recrystallized to granular textures. Mineral modes in the epidiosites range as follows: epidote (47–63%), quartz (31–40%), chlorite (0–11%), magnetite + sphene (4–5%).

The distribution of the two greenschist alteration styles is shown in Fig. 1a. Isochemical greenschist (type 1) assemblages occur over much of the western Solea graben, whereas epidiosite (type 2) assemblages form distinct zones. Partly epidiositized dykes are uncommon, and the paucity of such transitional assemblages suggests that epidiosite alteration was strongly controlled by fluid infiltration²³. Epidiosite zones occur immediately beneath argillic alteration zones in both dykes and lavas. Argillic alteration is texturally similar to epidiosite alteration, but is characterized by epidote-free mineral assemblages. Recent research drilling through Cyprus massive sulphide deposits confirms that similar argillic alteration immediately underlies the stockwork zones of the orebodies²⁴.

Figure 1a shows the location of three epidiosite zones within the western Solea graben, separated from one another by normal variance greenschist facies rocks. A 25-km², northeastward-trending zone within the middle to lower part of the Sheeted Dyke Complex and a smaller zone that apparently underlies the Apliki orebody, are characterized by whole-rock $\delta^{18}\text{O}$ values between +2.8 and +5.1 ‰ (Fig. 1b). A third zone of epidiosites occurs along the detachment near Lemithou (Fig. 1a), which has whole-rock $\delta^{18}\text{O}$ values up to +7.8‰. Isochemically metamorphosed greenschist to subgreenschist facies dykes range from +5.1 to +13.6‰ (Table 1). A north-west-south-east cross-section through the main epidiosite body (Fig. 1b, c) demonstrates the steep orientation of this zone, as reflected by nearly vertical contours in whole-rock $\delta^{18}\text{O}$. Away from the two shallow epidiosite zones, the regional oxygen isotope surfaces over the area of Fig. 1 are nearly flat-lying and are broadly conformable with the ophiolite pseudostratigraphy.

The complete recrystallization of rocks with diabasic igneous textures to granular hydrothermal aggregates and the develop-

ment of high-variance hydrothermal mineral assemblages imply that the epidiosite alteration is a metasomatic phenomenon involving large amounts of fluid²³. Whole-rock trace element and fluid inclusion/oxygen-isotope data for hydrothermal quartz from epidiosites provide corroboration. Two-phase, aqueous fluid inclusions from quartz homogenize to liquid at temperatures between 269 and 343 °C (Table 2). If the ocean depth above the Troodos complex was approximately 2.5 km¹⁸ and the epidiosites crystallized at depths of 1–2 km below the ocean-floor, the appropriate pressure-corrected²⁵ temperatures range from approximately 310–375 °C. When used with the quartz $\delta^{18}\text{O}$ data, these temperatures imply that epidiosite-forming fluids had $\delta^{18}\text{O}$ values near 0 ‰, close to the composition of unmodified Upper Cretaceous seawater. Freezing point depressions of fluid inclusions in quartz also imply near-seawater salinities (that is 3.5 wt% NaCl equivalent) for the fluids²⁵. Concentrations of many trace elements in the epidiosites depart markedly from those of other Solea dykes and lavas (Table 1). Epidiosites exhibit two-fold enrichments in Sr and order of magnitude depletions in Cu, Zn and Zr relative to all other Solea graben rocks of apparently equivalent primary composition, whereas Rb concentrations are very low in all dykes of the Solea graben (Table 1).

Transitions from epidiosite through argillite to the stockwork zones beneath massive sulphide deposits occur within the sheeted dykes of the western Solea graben. Thus, field relations imply that the epidiosites represent deep parts of a large, fossil submarine hydrothermal system. Inferred fluid temperatures (310–375 °C) overlap the range reported for active, high-temperature hydrothermal vents² and are similar to temperatures of hydrothermal alteration within stockworks subjacent to Troodos massive sulphides²⁶.

According to theoretical models of basalt/seawater interaction²⁷, epidote-quartz-chlorite assemblages form at fluid:rock ratios between 10 and 20. The extreme depletion of Zr (Table 1), an element generally considered to be 'immobile' during

Table 2 Oxygen isotope, fluid inclusion and mineralogic data for epidiosites of the Solea graben, Troodos ophiolite

Sample number	Epidote Fe*	Chlorite Mg†	Mean T_h (°C)‡	$\delta^{18}\text{O}$	(% versus SMOW)	
				Whole-rock	Quartz	Fluid§
7158403A	22–24	57–60	323 (24; 8)	2.8	5.8	+0.5
7158405	21–26	NP	343 (04; 11)	5.0	6.3	+1.5
7158414	23–25	48–58	269 (20; 11)	3.1	5.3	–1.2
80103	18–23	53–56	301 (04; 15)	3.5	6.6	–1.3
80104	21–26	57–64	295 (13; 19)	3.5	8.1	+2.8

* Octahedral (Fe/Fe + Al) × 100.

† (Mg/Mg + Fe) × 100.

‡ Mean homogenization temperatures (see text for discussion on pressure-corrected temperatures); one sigma absolute and number of determinations in parentheses.

§ Calculated with pressure-corrected²⁵ mean T_h , $\delta^{18}\text{O}$ for quartz, and quartz-H₂O O-isotope fractionations³¹.

NP, not present. SMOW, standard mean ocean water.

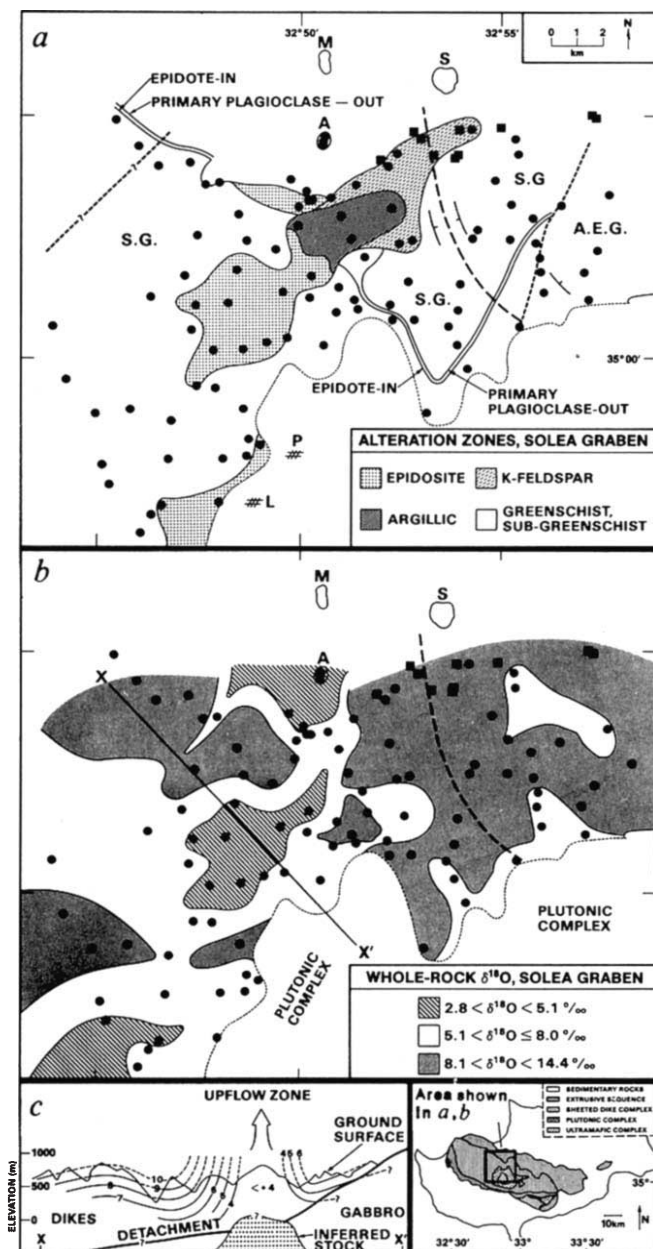


Fig. 1 Generalized structural, mineralogical and oxygen-isotope relations of the Solea graben area, Cyprus. Inset (bottom) shows map area of *a* and *b*, where dots (dykes) and squares (lavas) are sample localities and massive sulphide orebodies are: A, Apliki; M, Mavrovouni; S, Skouriotissa. Also shown in *a* and *b* are the Solea spreading axis (bold dashed line) and graben boundaries (light dashed lines), the trace of the dyke/gabbro detached contact (dotted line), and the distribution of hydrothermal assemblages (*a*) and whole-rock $\delta^{18}O$ values (*b*). Place names: A.E.G., Ayios Epiphanios graben; S.G., Solea graben; L, Lemithou; P, Pedhoulas. Part *c* is cross-section X-X' (trace shown in *b*), where $\delta^{18}O$ contours steepen as the main Solea epidosite/upflow zone is approached.

hydrothermal alteration²⁸, similarly implies a high fluid:rock ratio, metasomatic origin for the epidiosites. Indeed, extreme depletions of Cu and Zn in the epidiosites suggest that the high-temperature alteration of dykes has contributed ore constituents¹⁸ to overlying (now eroded) massive sulfide orebodies.

The distribution (Fig. 1a) and structure (Fig. 1c) of the epidiosite zones are not related in any obvious way to the dominant north-northwestward-trending spreading/structural grain of the

Solea graben, as defined by the orientation of dykes and the graben axis. Sheeted dykes of the Solea graben were originally emplaced vertically within the spreading axis and were subsequently transported laterally and tilted by extension-related faulting. If the epidiosite zones were closely related to spreading within the Solea axis and were subsequently deformed by tectonic extension, the epidiosites would be expected to have a north-northwestward-striking, northwestward-dipping orientation (that is, parallel to the axis of the Solea graben). The north-east strike and nearly vertical attitude of the main epidiosite zone suggest that its development is unrelated to the Solea spreading axis and argues for a post-graben (that is, post-tectonic extension) origin. Extreme tectonic attenuation of the crust was apparently the terminal phase in the spreading history of the Solea axis³⁰. Only a few examples of steep, unrotated dykes cutting deformed, highly rotated dykes have been observed within the graben, indicating that most of the magmatic spreading by dyke injection preceded structural attenuation. Thus, the epidiosite alteration appears to post-date both magmatic and structural spreading in the Solea graben, but may reflect renewed magmatism in the Solea region immediately before a ridge jump approximately 8 km east of the Solea axis¹.

If the ore deposits and epidiosites in the Solea graben formed in an off-axis setting, there must have been a related magmatic event capable of driving hydrothermal circulation. The outcrop patterns of upwelling zones within the dyke complex may reflect the distribution of off-axis, hypabyssal intrusions. Although the inferred off-axis intrusions are largely unexposed, we have found one small gabbro stock north-west of Pedhoulas (Fig. 1a) that is relatively unaltered and cuts highly rotated epidiosite dykes. Our data suggest that the establishment of long-lived, stationary hydrothermal cells and the formation of significant sulphide accumulations at the sea floor may be promoted by the intrusion of shallow stocks into previously deformed and tectonically thinned oceanic crust.

We thank Carol Eddy, Steve Hurst and Todd Ramsden for assistance and helpful discussion, and David Mass for performing many of the isotope analyses. The comments of E. T. C. Spooner and an anonymous reviewer are appreciated. Supported by NSF (grants EAR83-0659 and 86-10503) and the Science and Technology Division of Unocal Corporation.

Received 7 July; accepted 22 October 1986.

- Varga, R. J. & Moores, E. M. *Geology* **13**, 846-850 (1985).
- Haymon, R. M. & MacDonald, K. C. *Am. Sci.* **73**, 441-449 (1985).
- Hekinian, R. *et al. Science* **219**, 1321-1324 (1983).
- MacDonald, K. C. *A. Rev. Earth planet. Sci. Lett.* **10**, 155-190 (1982).
- Normark, W. R. *et al. Geology* **11**, 158-163 (1983).
- Spiess, F. N. *et al. Science* **207**, 1421-1433 (1980).
- Alabaster, T. & Pearce, J. A. *Econ. Geol.* **80**, 1-16 (1985).
- Cocker, J. D. *et al. Earth planet. Sci. Lett.* **61**, 112-122 (1982).
- Coish, R. A. *Contr. Miner. Petr.* **60**, 255-270 (1977).
- Elthon, D. & Stern, C. *Geology* **6**, 464-468 (1978).
- Geary, E. E. & Kay, R. W. in *Tectonic and Geologic Evolution of Southeast Asian Seas and Islands*, Part 2, 139-156 (Am. Geophys. Union) Geophysical Monograph Series 27 (1982).
- Gregory, R. T. & Taylor, H. P. Jr *J. geophys. Res.* **86**, 2737-2755 (1981).
- Schiffman, P. *et al. Earth planet. Sci. Lett.* **70**, 207-220 (1984).
- Gass, I. G. & Smewing, J. D. *Nature* **242**, 26-29 (1973).
- Heaton, T. H. E. & Sheppard, S. M. F. in *Volcanic Processes and Ore Genesis*, 42-57 (Geological Society of London, 1966).
- Spooner, E. T. C. *et al. Contr. Miner. Petr.* **47**, 41-62 (1974).
- Moores, E. M. *et al. Geology* **12**, 500-503 (1984).
- Spooner, E. T. C. *Geol. Ass. Can. spec. Pap.* **20**, 685-704 (1980).
- Oudin, E. & Constantinou, G. *Nature* **308**, 349-353 (1984).
- Spooner, E. T. C. in *Volcanic Processes and Ore Genesis*, 58-71 (Geological Society of London, 1977).
- Schiffman, P. *et al. Eos* **66**, 402 (1985).
- Ishizuka, H. J. *Petrol.* **26**, 391-417 (1985).
- Wilson, R. A. M. *The Geology of the Xeros - Troodos Area* (Cyprus Geological Survey Memoir 1, 1959).
- International Crustal Research Drilling Program *Geotimes* **29**, 12-14 (1984).
- Potter, R. W. *J. Res. U.S. geol. Surv.* **5**, 603-607 (1977).
- Schiffman, P. *et al. Eos* **66**, 1128 (1985).
- Spooner, E. T. C. & Bray, C. T. *Nature* **266**, 808-812 (1977).
- Reed, M. H. *Econ. Geol.* **78**, 446-485 (1983).
- Pearce, J. A. & Cann, J. R. *Earth planet. Sci. Lett.* **19**, 290-300 (1973).
- Varga, R. J. & Moores, E. M. *Geol. Soc. Am. Abstr. Progr.* **18**, 194 (1986).
- Clayton, R. N. *et al. J. Geophys. Res.* **77**, 3057-3067 (1972).

The effect of sparse vegetation on the transport of dune sand by wind

Ralf Buckley

AMDEL, Flemington Street, Frewville, SA 5063, Australia

The precise effect of a sparse plant cover on the transport of dune sand by wind is a critical factor in understanding the morphology, evolution and global distribution of aeolian dunes¹⁻³. It is also of considerable practical importance in sand stabilization, rehabilitation and restoration ecology. Here I report the first detailed experimental quantification of this effect using living plants in patterns approximating natural vegetation. For wind velocities V (at 0.5 m elevation) up to 15 m s^{-1} and plant cover C up to 17%, the rate of aeolian sand transport, $q \text{ g cm}^{-1} \text{ s}^{-1}$, is closely approximated by $q = B[V(1 - kC) - V_t]^3$, where B is Bagnold's constant⁴, V_t is a threshold velocity, and k is a constant dependent on plant geometry. For typical small erect or spreading herbaceous dune plants, $k = 0.018$; for small rounded stemless plants $k = 0.046$.

This relationship was determined in a large sand-driving wind tunnel⁵, using a well-sorted siliceous aeolian dune sand with mean diameter 0.15 mm. The tunnel is 38 m long overall, with directional vanes and a large vertical-plane expansion chamber to remove rotational components in the airflow, and a long upstream section to stabilize the longitudinal velocity component and minimize cross-stream turbulent components. The test section is 16 m long, 1 m wide and 0.6 m high. Three plant types were used, corresponding to characteristic shapes of common small herbaceous dune plants; erect and spreading (type F), 10 cm high and 12 cm across; graminiform (type G), 12 cm high and 9 cm across; and rounded (type R), 5 cm high and 7 cm across. The plants were mounted in narrow copper tubes soldered to moveable baseplates under the same bed. Plants were distributed in 7 contiguous 1 m squares in the central segment of the test section, with the same density, n , in each square. Densities of 0, 1, 2, 3, 4, 5, 7, 10 and 15 plants per square metre were tested for each plant type, representing projected cover C of 0–17% for type F, 0–10% for type G and 0–3.8% for type R. Within each 1 m × 1 m square, the position of each individual plant was assigned by using two independent random numbers from a uniform random distribution on [0, 1] as its coordinates. Sand transport was measured with three high-precision vertical profile traps 50 cm downstream of the vegetated area: one in the centre of the tunnel, and one 20 cm in from each edge. Each trap consists of a knife-edged vertical slot 1 cm across, divided into 1 cm square orifices leading to independent collectors under the tunnel.

Particular precautions were taken to minimize experimental error: the profile traps were not opened until airflow and ripple bedform had stabilized; the sand bed was deep enough to cover the moveable baseplates and mounting tubes completely, and was levelled before each experiment; air velocity profiles were measured both upstream and downstream of the experimental section, using banks of pitot-tube anemometers; and the tests were replicated several times, using different random plant layouts, for each windspeed, plant type and density: several hundred replications in all. The total error from measurement imprecision is less than 1%. Other factors which must be taken into account are as follows. The wind velocity gradient from the edge to the centre of the tunnel increases slightly along the experimental section, particularly at high velocities, so that the effective velocity of the sand-moving air current may not be constant along the test section. There is also an edge-to-centre horizontal velocity gradient, particularly on bare sand: tests using a row of 10 profile traps showed that two traps 20 cm from the margins consistently overestimated the 10-trap mean

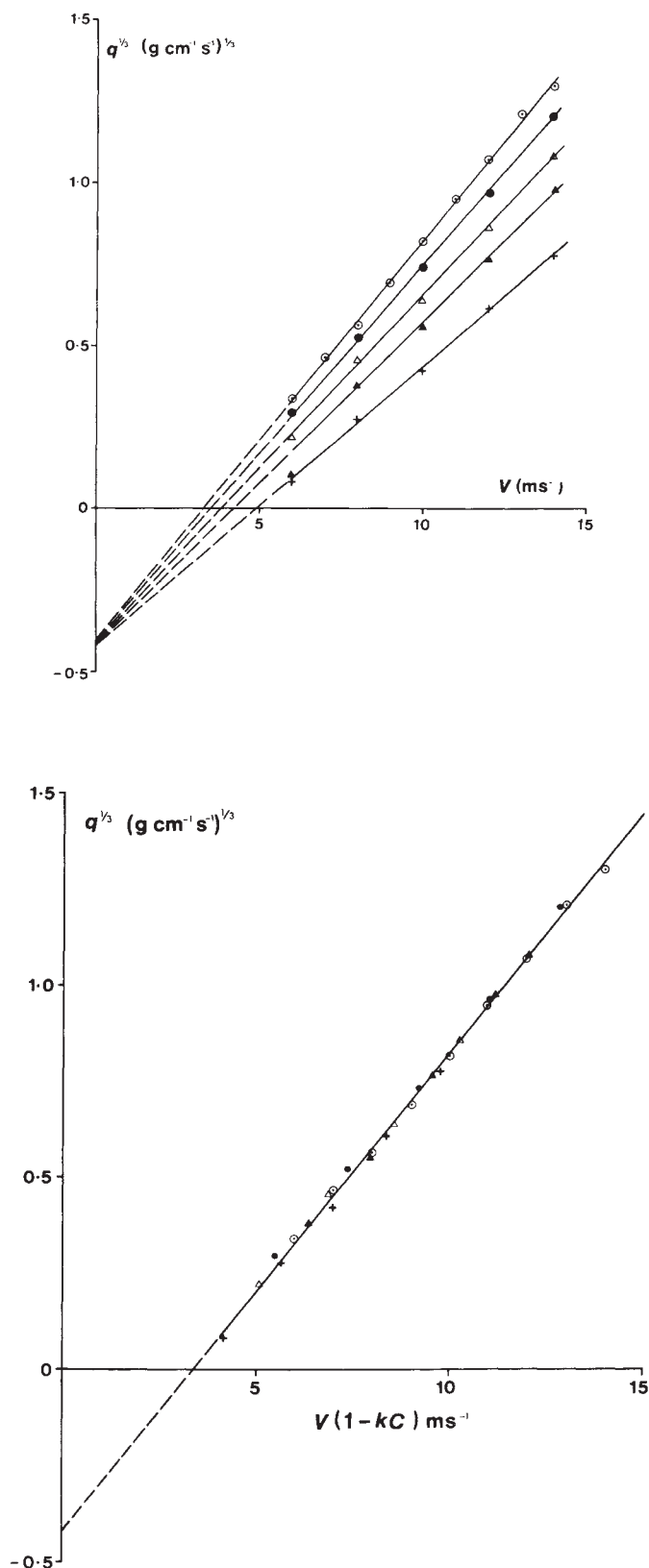


Fig. 1 *a*, Relation of cube root of sand transport rate, $q^{1/3}$, to wind velocity, V , at densities $n = 0$ –15 plants per m^2 of a small erect spreading herbaceous dune plant. Percent projected cover, C , is given by $C = 1.13 n$ for this plant species. $\circ$, $n = 0$; $\bullet$, $n = 4$; $\triangle$, $n = 7$; $\blacktriangle$, $n = 10$; $+$, $n = 15$. *b*, Overall relation between $q^{1/3}$ and $V(1 - kC)$ for all n and V . Slope and regression of overall regression line agree with predictions from individual regressions in Fig. 1*a* to within <1%. Slope of overall line is $B^{1/3}$ where B is Bagnold's⁴ coefficient.

by approximately 1.5% for low wind velocities, and underestimated it by 5% for higher velocities, when the 'central jet' effect is more pronounced. The sand flow rate at each trap is also influenced by the precise lateral position of the individual plants at the downstream end of the vegetated section: this effect, estimated by comparing the results from the two edge traps, is the principal source of stochastic variation at low plant densities, but insignificant at higher densities. Finally, sand transport may be affected by the precise positions of the individual plants: the magnitude of such variation was determined by replication with different random patterns. The overall accuracy of the final data from each individual replication, allowing for all these sources of error and variation, ranges from $\pm 5\%$ for the higher plant densities to $\pm 10\%$ for the lower. The good fit of the aggregate data to the straight lines in Fig. 1 confirms the success of these measures to reduce experimental error.

Results for plant type F are shown in Fig. 1. For each plant density n , the regression of $q^{1/3}$ on V has $R^2 > 0.998$ (excluding the point for $n = 10$, $V = 6$) (Fig. 1a). The general equation for the straight lines in Fig. 1a is

$$q^{1/3} = Vf_1(C) - f_2(C) \quad (1)$$

For $C = 0$, bare sand, $f_1(C)$ should be equal to the cube root of Bagnold's⁴ factor B , defined as $B = [0.174/\log(z/k')]^3 C_s \sqrt{(d/D) \rho/g}$: where z is the height in cm at which V (in cm s^{-1}) is measured; k' is approximately 1 cm; C_s is a sorting factor ranging from 1.5–2.8 and equal to 1.8 for naturally sorted dune sand; d is mean grain diameter, and D is a standard grain diameter of 0.25 mm; ρ is the density of air, and g is the gravitational constant. In fact $f_1(0) = 1.23 \times 10^{-3}$, which is within 1% of the calculated value for $B^{1/3}$: less than the uncertainty in C_s , d or k' . The regression slopes, $f_1(C)$ in equation (1), may be approximated closely by $f_1(C) = B^{1/3}(1 - kC)$, where $k = 0.018 \pm 0.001$ ($R^2 = 0.99$). The intercept $f_2(C)$ is approximately constant, equal to -0.41 ± 0.01 . Hence

$$q = B[V(1 - kC) - V_t]^3 \quad (2)$$

where V_t is a threshold velocity of $3.33 \pm 0.08 \text{ m s}^{-1}$. Note that this is the threshold velocity for bare sand of mean diameter 0.15 mm; on vegetated sand, the threshold velocity for sand movement is $V_t/(1 - kC)$, increasing with C . For $C = 17\%$, for example, it is 4.8 m s^{-1} .

Similar equations describe the results for the other two plant types. The coefficients for types G and R were estimated less precisely than for type F since the degree of replication was less. For plant type G, with approximately the same horizontal projected area as type F, k and V_t are statistically indistinguishable from the corresponding values for type F. For type R, $k = 0.046 \pm 0.012$, statistically distinct from k for type F at $P < 0.05$: a given cover of small rounded plants reduces sand transport more than an equal cover of erect spreading plants, as might be anticipated. The intercept term, $f_2(C)$, increases slightly with C for plant types G and R, but the variation is not statistically significant. Hence the equation $q = B[V(1 - kC) - V_t]^3$ applies to all three types, with only k dependent on plant geometry.

This remarkably simple relation between q , V and C implies that for constant V , q is a cubic polynomial function of C : for $V = 15 \text{ m s}^{-1}$ and plant type F, for example, $q = 3.017 - 0.209C + 0.00483C^2 - 0.0000372C^3$. Even a very sparse plant cover can cause a significant reduction in q (Table 1). It should be emphasised that this relation is only tested for $C = 0$ –17%. If it holds for $C > 17\%$, it implies that for any given V there is a cutoff value⁶ of C , C_{cut} , at which q falls to 0; and that as $C \rightarrow C_{\text{cut}}$, q is very small (Table 1). For $C > C_{\text{cut}}$ equation (2) predicts $q < 0$. This might represent sand accumulation in densely vegetated areas. Alternatively, equation (2) may not describe the q – V – C relation adequately for higher C .

Table 1 Effect of plant cover on sand transport rate at constant wind velocity (from equation 2)

		Sand transport rate as per cent of bare-sand rate for wind velocity:	
% Cover, C of plant type F		$V = 10 \text{ m s}^{-1}$	$V = 15 \text{ m s}^{-1}$
Within range of C tested experimentally	0	100	100
	1	92	93
	2	85	87
	3	78	81
	4	71	75
	5	65	69
	8	48	54
	9	43	50
	17	16	22
	20	10	16
Extrapolated for C outside tested range	23	5	10
	30	0.7	3
	34	0.06	1.0
	37	0	0.3
	43	[−0.4]	0

The closest previous approach to this investigation is that of Wasson *et al.*^{6,7}. They measured the horizontal distribution of wind velocities at 1 m elevation around individual dune shrubs in the field⁷, used Bagnold's equation⁴ to calculate the horizontal distribution of sand transport rates around individual shrubs, summed these rates to yield aggregate q at various C , and fitted curves⁶ to plots of aggregate q against C . They did not measure q directly; and they assumed that Bagnold's equation was applicable at a plant-size scale, whereas it is possible that the wake effects of plants, notably turbulence and acceleration/deceleration, modify air velocity and sand transport profiles so that Bagnold's equation no longer applies. In addition, they fitted a relatively complex exponential function to their data:

$$q = B[Vf(C) - V_t]^3 \quad (3)$$

where $f(C) = \exp(-k_1 C - k_2 C^2)$, with $k_1 = 6.17 \times 10^{-3}$ and $k_2 = 4.29 \times 10^{-4}$. They did not provide confidence limits for the curves or coefficients. Despite these limitations, their predictions (equation (3)) are not dissimilar to those of equation (2). The comparison requires standardization of the height, z cm, at which V is measured, using the relation $(V_{z_1} - V_t)/(V_{z_2} - V_t) = \log(z_1/k')/\log(z_2/k')$. For C in the range 0–17%, $\exp(-k_1 C - k_2 C^2)$ ranges from 1 to 0.795, and $(1 - 0.018 C)$ from 1 to 0.694. For $V_t = 3.33 \text{ m s}^{-1}$ and V measured at 1 m elevation, equation (3) predicts $C_{\text{cut}} = 44\%$ for $V = 10 \text{ m s}^{-1}$ and 52.5% for $V = 15 \text{ m s}^{-1}$; corresponding values from equation (2) are 35% and 41% respectively.

Besides the differences in the range of C considered, the difference between the results presented here and those of Wasson *et al.*^{6,7} may well be because of differences in plant geometry and layout. Neither Wasson *et al.*^{6,7} nor the present study attempted to incorporate these factors explicitly into the q – V – C equation. There have, however, been a number of attempts to incorporate the effects of non-erodible roughness elements into soil erosion equations used in agricultural engineering contexts^{8,9}. These agricultural studies are not directly applicable to the sparse vegetation of aeolian desert dunes, because they have generally been in temperate or semi-arid regions where moisture is a significant control on soil movement, and have generally been concerned with dense rather than open plant cover¹⁰, and with crop residues^{11,12}, mulches¹³, windbreaks or barriers¹⁴, or (in windtunnel studies) with scale models or simulated or artificial crops¹⁵, rather than natural living vegetation. The most detailed equation, proposed by Lyles and Allison¹⁵ for areas covered by standing stubble, incorporated silhouette area (horizontal projected area) and plant

spacing both parallel and perpendicular to wind direction, as well as plant density. For natural dune vegetation, measures of flexibility and profile shape will probably also be necessary. In particular, plants with the same vertical and horizontal projected areas may have very different effects on airflow and sand transport if one grows as a short-stemmed sand-filled mound, with the main silhouette area close to the sand surface, and the other as a long-stemmed shrub, with the main silhouette area well above the sand surface. The effects of a mixed plant cover including perennial shrubs, mound-forming species and short-lived herbaceous species will also need to be considered, as will the effects of fluctuations in wind velocity. Direct experimental measurement in the field would be the simplest and most reliable approach. There is no readily available means for measuring sand transport over a sufficiently short time period to relate it to the constantly fluctuating wind velocity in the field, but instruments using piezoelectric effects or β -ray attenuation¹⁶ may achieve this soon. Field measurements will also be needed to examine sand transport on sloping surfaces such as dune flanks.

I acknowledge the assistance of the Australian Academy of

Sciences and Academia Sinica. The experimental work was carried out at the Institute of Desert Research in Lanzhou, China, with the assistance of Dr Ling Yuquan and approval of Professor Zhu Zhenda, Director. Presentation of the results benefitted considerably from discussions at the International Sedimentological Congress in August 1986, and from comments by referees and Dr L. Lyles and Dr D. W. Fryrear of USDA.

Received 18 September 1986; accepted 3 December 1986.

1. Wasson, R. J. & Hyde, R. *Nature* **304**, 337–339 (1983).
2. Wasson, R. J. *Nature* **309**, 92 (1984).
3. Mabbitt, J. A. *Nature* **309**, 92 (1984).
4. Bagnold, R. A. *The Physics of Blown Sand and Desert Dunes* (Chapman & Hall, London, 1941).
5. Ling, Y. Q. *Acta geographica sinica* **35**, 174–183 (1980).
6. Wasson, R. J. & Nanninga, P. M. *Earth Surf. Proc. Landf.* (in the press).
7. Ash, J. E. & Wasson, R. J. *Zeitschr. Geomorph. Suppl.* **45**, 7–25 (1983).
8. Lyles, L., Hagen, L. J. & Skidmore, E. L. *Agron. Monog.* **23**, 177–188 (1983).
9. Fryrear, D. W. 1985 *Trans. Am. Soc. Agric. Eng.* **28**, 781–784 (1985).
10. Marshall, J. K. *Proc. Int. Grassl. Congr.* **11**, 19–24 (1970).
11. Lyles, L. & Allison, B. E. *Trans. Am. Soc. Agric. Eng.* **24**, 405–408 (1981).
12. Bilbro, J. D. & Fryrear, D. W. *J. Soil Water Cons.* **40**, 358–360 (1985).
13. Armbrust, D. V. *Trans. Am. Soc. Agric. Eng.* **20**, 904–905 (1977).
14. Aase, J. K., Siddoway, F. H. & Black, A. L. *J. Soil Water Cons.* **40**, 354–357 (1985).
15. Lyles, L. & Allison, B. E. *Trans. Am. Soc. Agric. Eng.* **19**, 61–64 (1976).
16. Thompson, E. M., Treafitt, H. N. & Tomb, T. F. 1984. *U.S. Dept. Labour Rep.* IR1148 (1984).

Evidence for a relationship between hydrocarbons and authigenic magnetite

R. D. Elmore*, M. H. Engel*, L. Crawford*, K. Nick*, S. Imbus* & Z. Sofer†

* School of Geology and Geophysics, The University of Oklahoma, Norman, Oklahoma 73019, USA

† Cities Service Research, Box 3908, Tulsa, Oklahoma 74102, USA

Establishing a relationship between hydrocarbon migration and the precipitation of authigenic magnetite in sedimentary rocks is of significant interest with respect to (1) elucidating mechanisms for remagnetization and establishing the origin of secondary magnetizations residing in magnetite, (2) developing a method to date hydrocarbon migration events by determining the time of remanence acquisition by palaeomagnetic methods, and (3) evaluating whether searching for anomalous concentrations of diagenetic magnetic minerals and/or aeromagnetic anomalies is justified as a relatively inexpensive exploration tool. The direct association of hydrocarbon migration and the precipitation of authigenic magnetite has, however, not been established. In this paper we report palaeomagnetic, rock magnetic, petrographic and geochemical results of a study of samples from Permian speleothems and gilsonite found in the Ordovician Arbuckle Group in southern Oklahoma. The results indicate that there is a genetic relationship between hydrocarbon migration and the precipitation of authigenic magnetite.

It has recently been speculated that the precipitation of authigenic magnetite in sedimentary units may in some cases be related to the migration of hydrocarbons and/or associated brines through the host rocks. For example, finely dispersed magnetite has been reported from oil-saturated sedimentary rocks in Russia¹. Donovan *et al.*² reported magnetite in altered Permian clastic rocks above the producing horizons at the Cement oil field, Oklahoma, and hypothesized that hydrocarbon-related brines migrated from depth upward along faults, causing reduction of haematite and iron hydroxides, and the formation of magnetite. They also suggested that the magnetite is the source of an aeromagnetic anomaly over the Cement oil field. Reynolds *et al.*³ noted that the Cement oil field magnetite reported by Donovan *et al.*² exhibits synthetic textures, and suggested that it represents contamination from drilling. Reynolds *et al.*^{3,4} did, however, report the occurrence of ferrimagnetic

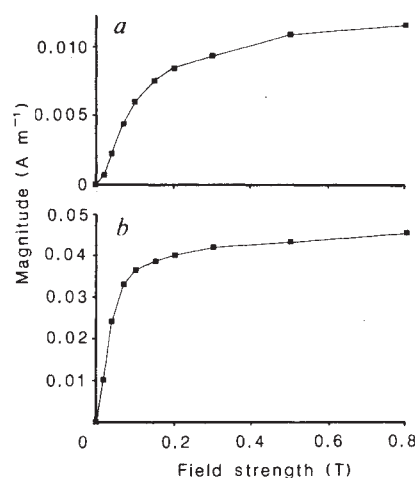


Fig. 1 Representative IRM acquisition curves for specimens from the gilsonite (a) and the hydrocarbon-saturated speleothems (b). Both curves show a steep initial rise in IRM up to 0.2–0.3 T followed by a flattening-off of the slope. This suggests a low-coercivity phase such as magnetite is the dominant phase, although the gradual rise above 0.3 T suggests that a higher-coercivity phase such as haematite is also present.

pyrrhotite from the Cement field and suggested that it precipitated as a result of hydrocarbon-related fluids.

Chemical remanent magnetizations (CRMs) that reside in diagenetic magnetite have been previously reported for several units (for example, refs 5, 6). Evidence cited includes secondary magnetizations acquired long after deposition as well as the extraction and identification of spherical authigenic magnetite from the units. Although some secondary magnetizations residing in authigenic magnetite are probably thermoviscous in origin (for example, ref. 7), independent palaeotemperature information, in conjunction with theoretical curves that relate blocking temperatures to relaxation time⁸, can often be used to rule out a thermoviscous origin for some of the magnetizations. The chemical origin and mechanism(s) for precipitation of the diagenetic magnetite that carries a CRM are unknown, although a relationship with hydrocarbon-related fluids⁵ or tectonic fluids has been suggested⁹.

Previous investigations of the Lower Ordovician Arbuckle Group in southern Oklahoma revealed that the rocks contain a late Palaeozoic CRM residing in haematite^{10,11}. In rocks that contained hydrocarbons, however, the magnetization was

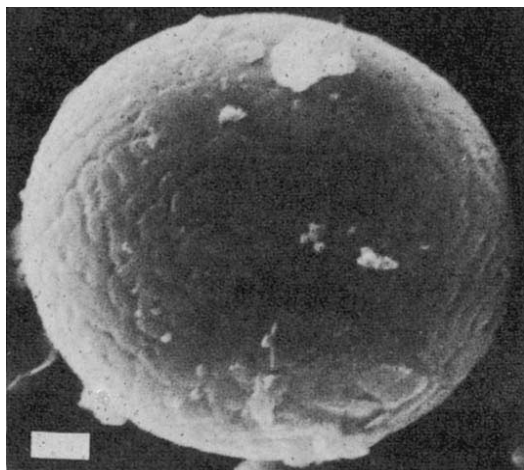


Fig. 2 Magnetic sphere extracted from the gilsonite and identified as magnetite. Similar spheres are found in the extracts from the speleothems. Scale bar, 2 μm .

observed to be a late Palaeozoic CRM residing in magnetite¹². In the present study, we analysed samples of hydrocarbon-saturated calcite speleothems and gilsonite for their organic compositions and remanent magnetization. The speleothems occur as flowstones and were collected from caves and fissures in Arbuckle Group limestones north of the Wichita Mountains. Amphibian and reptilian fossils found interbedded with the flowstones indicate that the deposits are Permian in age (Louis Jacobs, Southern Methodist University, personal communication). The speleothems are mostly composed of clear to dark calcite bands, the dark bands containing abundant hydrocarbon inclusions. Preliminary petrographic study, including ultraviolet fluorescence microscopy, indicate that the hydrocarbons are found as abundant small ($<3\ \mu\text{m}$) hydrocarbon inclusions along growth planes, larger (commonly 25–100 μm) hydrocarbon fluid inclusions within crystals, and as inclusions associated with fractures and crystal boundaries. The inclusions along growth bands are thought to be primary. Samples of gilsonite (black, shiny asphaltite with conchoidal fracture) were collected from an active hydrocarbon seep in Arbuckle Group limestones in a quarry in the Arbuckle Mountains in southern Oklahoma.

Samples from the calcite speleothem deposits were collected with a portable coring apparatus and oriented *in situ* with a clinometer and Brunton compass. Standard palaeomagnetic specimens were cut from the samples and their natural remanent magnetization was measured with a United Scientific three-axis cryogenic magnetometer. The specimens were demagnetized by alternating field techniques in a shielded two-axis tumbling unit and by thermal techniques in a Schonstedt TSD-1 unit. Direction and intensity results were plotted on orthogonal projections¹³ and the direction and magnitude of components removed during demagnetization were calculated using a principal-component analysis program. Specimens of the speleothems and the gilsonite were also subjected to isothermal remanent magnetization (IRM) experiments using a water-cooled electromagnet to aid in identification of the magnetic minerals.

Extreme care was taken during collection and subsequent magnetic extraction of the gilsonite and speleothems to reduce the possibility of contamination. The gilsonite was dissolved in toluene, and the solution was centrifuged to separate the mineral matter from the bulk of the organic material. The slurry containing the mineral material was then run through a magnetic extraction line to separate the magnetic and non-magnetic components. Samples of the speleothem were dissolved in glacial acetic acid buffered to $\text{pH} = 4$. The residue was also run through a magnetic extraction line. All extracts were examined by X-ray diffraction (XRD) and scanning electron microscope-energy

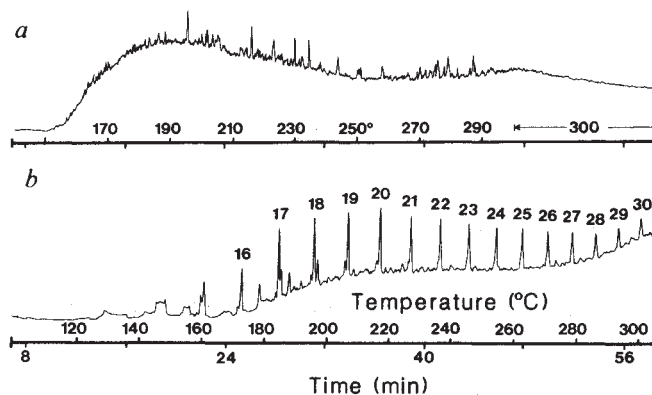


Fig. 3 Chromatograms of the saturated fractions from the gilsonite (a) and hydrocarbon-saturated speleothems (b). The gilsonite analyses were done with a Hewlett-Packard 5890A GC equipped with a flame ionization detector and a 30 m $\times$ 0.25 mm i.d. J&W bonded-phase column coated with a 1.0 μm film of DB-5. The injector and detector temperatures were 300 $^{\circ}\text{C}$. Sample injection was done in the splitless mode. The carrier gas (He) flow rate was 1.0 ml min^{-1} . The oven-temperature programme was 40 $^{\circ}\text{C}$ isothermal for 4 min, then 40 $^{\circ}\text{C}$ to 150 $^{\circ}\text{C}$ at 20 $^{\circ}\text{C min}^{-1}$ and 150 $^{\circ}\text{C}$ to 300 $^{\circ}\text{C}$ at 4 $^{\circ}\text{C min}^{-1}$. The run was finished isothermally at 300 $^{\circ}\text{C}$. Analyses of the hydrocarbon-saturated speleothems were performed with a Varian 3300 GC equipped with an FID detector and a column identical to that described above. The injector and detector temperatures were 300 $^{\circ}\text{C}$. Sample injection was done in the splitless mode. After 0.5 min the split was adjusted to 50:1. The carrier gas (He) flow rate was 2 ml min^{-1} . The oven-temperature programme was 40 $^{\circ}\text{C}$ isothermal for 1 min, then 40 $^{\circ}\text{C}$ to 100 $^{\circ}\text{C}$ at 10 $^{\circ}\text{C min}^{-1}$ and 100 $^{\circ}\text{C}$ to 300 $^{\circ}\text{C}$ at 4 $^{\circ}\text{C min}^{-1}$. The run was finished isothermally at 300 $^{\circ}\text{C}$.

dispersive analysis (SEM-EDA) techniques. Thin sections of the speleothems, and polished grain mounts of the extracts, were examined under both transmitted and reflected light.

The gilsonite and organic material present in the dark bands of the calcite speleothems were also analysed for their respective saturate hydrocarbon compositions by gas chromatography (GC). Portions of the light and dark calcite bands of the speleothem were also ground to fine powders ($\sim 75\ \mu\text{m}$), exhaustively extracted with CH_2Cl_2 and boiled in H_2O_2 to remove organic matter. Portions of the purified calcite were subsequently digested with phosphoric acid and analysed for their respective stable carbon and oxygen isotopic compositions.

IRM acquisition curves for specimens of the gilsonite suggest that the magnetization is dominated by a low-coercivity mineral such as magnetite although the presence of a higher-coercivity mineral such as haematite is also indicated (Fig. 1a). Several phases were found in the magnetic extracts of the gilsonite. Magnetic spheres less than 15 μm in diameter are common (Fig. 2). The XRD patterns obtained with $\text{FeK}\alpha$ radiation indicate the presence of magnetite in the extracts containing the spheres. EDA results, which indicate that iron is the only detectable element in the spheres, and examination of polished grain mounts in reflected light, are consistent with the interpretation that the spheres are magnetite. Botryoidal aggregates, with iron as the only element detected based on EDA, are also present. The spheres are similar to those reported by other workers and interpreted as authigenic magnetite (for example, ref. 5). The composition and form of these phases in the gilsonite strongly suggest they are also authigenic magnetite.

Platy aggregates similar to authigenic haematite are also present. Iron was the only element detected in these phases by EDA, and haematite was identified based on the XRD patterns from extracts dominated by the platy aggregates. Non-magnetic phases found in the gilsonite include authigenic pyrite, calcite and kaolinite.

Figure 3a shows a chromatogram of the saturate hydrocarbon fraction of the gilsonite. The chromatogram shows depletion of the *n*-alkanes, which is typical for degraded hydrocarbons. The occurrence of authigenic magnetite in apparently biodegraded hydrocarbons has been discussed previously^{12,14}.

The dark bands in the speleothems possess up to an order of magnitude stronger magnetization (average $8 \times 10^{-4} \text{ A m}^{-1}$) than the lighter bands (average $6 \times 10^{-5} \text{ A m}^{-1}$). Specimens of the dark calcite exhibit stable linear decay to the origin during both alternating field and thermal demagnetization. Commonly, over 95% of the magnetization was removed by 125 mT and maximum unblocking temperatures were below 580 °C. The dark calcites contain a CRM with a south-easterly and shallow 'Kiaman' or Permian direction of magnetization (declination = 159° E, inclination = 3°, $k = 29$, $\alpha_{95} = 5^\circ$, $n = 25$). The magnetization in the light bands, which do not contain as many hydrocarbon inclusions as the dark bands, was too weak to yield stable results during demagnetization.

IRM acquisition curves (Fig. 1b) for all specimens show that the magnetization is dominated by a low coercivity phase such as magnetite. Thermal demagnetization of the IRM shows that maximum unblocking temperatures are below 580 °C. These data, in conjunction with the demagnetization results, suggest that the magnetization mainly resides in magnetite.

The hydrocarbons extracted from the calcites are at most only slightly degraded (Fig. 3b). Magnetic extracts from the dark calcites contain spherical, botryoidal, and other authigenic forms which, based on EDA, contain iron as the only detectable element. The XRD patterns of extracts from the calcites also indicate the presence of magnetite. The botryoidal and spherical forms are similar to authigenic magnetite reported in the literature and found in the gilsonite.

The magnetization in the speleothems was acquired in the Permian, resides in authigenic magnetite, and is interpreted as a CRM. The fact that cave deposits form at relatively shallow depths and that the Arbuckle Group speleothems were never deeply buried probably eliminates the possibility of a thermoviscous origin for the magnetization. Stable carbon and oxygen isotope values for the light ($\delta^{13}\text{C}_{\text{PDB}} -6.7\%$; $\delta^{18}\text{O}_{\text{PDB}} -4.2\%$) and dark ($\delta^{13}\text{C}_{\text{PDB}} -6.2\%$; $\delta^{18}\text{O}_{\text{PDB}} -4.7\%$) speleothems are consistent with a shallow depth of formation from fresh water¹⁵. The similarity of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values for the adjacent dark and light calcites favours a common origin.

The presence of authigenic magnetite in the gilsonite indicates a relationship between hydrocarbons and magnetite. The results from the light and dark speleothems also suggest that there is a relationship between hydrocarbons and authigenic magnetite; the stronger magnetization invariably resides in the dark calcites, where the hydrocarbons became concentrated. The occurrence of hydrocarbon inclusions along growth planes suggests that hydrocarbons seeped into the caves during the precipitation of the speleothems and were trapped in the calcite crystals. This relationship between intensity of magnetization and abundance of hydrocarbons leads us to propose that chemical conditions created by the hydrocarbons could have caused precipitation of authigenic magnetite and acquisition of the associated CRM. The fact that the time of acquisition of remanence is consistent with the age of the speleothems also provides a test that hydrocarbons can cause acquisition of a stable secondary magnetization residing in magnetite.

Although the results from the Arbuckle Group indicate a spatial and temporal relationship between authigenic magnetite and hydrocarbons, the mechanism for magnetite formation is not yet known. The fact that the gilsonite is degraded and contains magnetite suggests that microbial attack of hydrocarbons is one possible diagenetic pathway that might contribute to the formation of authigenic magnetite. This connection, however, is only tentative at best; almost all gilsonite or similar bitumen found at the Earth's surface is degraded. The magnetite may have formed during biodegradation, but an origin related

to other diagenetic processes cannot be ruled out. The fact that the hydrocarbons in the speleothems are not extensively degraded also argues against biodegradation as the only mechanism for precipitation of authigenic magnetite. We are now investigating this problem by conducting laboratory simulation experiments.

Support for this research was provided by the donors of the Petroleum Research Fund, administered by the American Chemical Society (PRF # 18484-AC2) and a grant from ARCO to R.D.E., and by the NSF (EAR-8352055) to M.H.E. The Energy Resources Institute at the University of Oklahoma also provided partial support.

Note added in proof: Analyses of the speleothems in a newly constructed magnetically shielded room confirm the results described above, although stable linear decay to ~620 °C in some specimens indicates the presence of a weak component residing in haematite. Haematite has also been found in the magnetic extracts from the speleothems.

Received 31 July; accepted 4 November 1986.

1. Bagin, V. I. & Malumgan, L. M. *Izv. Akad. Nauk. SSSR, Fiz zeml.* 4, 73-79 (1976).
2. Donovan, T. J., Forgy, R. L., & Roberts, A. A. *Bull. Am. Ass. Petrol. Geol.* 63, 245-248 (1979).
3. Reynolds, R. L., Fishman, N. S., Hudson, N. R., Karachewski, J. A. & Goldhaber, M. B. *Eos* 66, 867 (1985).
4. Reynolds, R. L., Fishman, N. S., Grauch, R. I. & Karachewski, J. A. *Eos* 65, 866 (1984).
5. McCabe, C., Van der Voo, R., Peacor, D. R., Scotese, R. & Freeman, R. *Geology* 11, 221-223 (1983).
6. Wisniowiecke, M. J., Van der Voo, R., McCabe, C. & Kelly, W. C. *J. geophys. Res.* 88, 6540-6548 (1983).
7. Kent, D. V. *Geophys. Res. Lett.* 12, 805-808 (1985).
8. Middleton, M. F. & Schmidt, P. W. *J. geophys. Res.* 87, 5351-5359 (1982).
9. Oliver, J. *Geology* 14, 99-102 (1986).
10. Elmore, R. D., Dunn, W. & Peck, C. *Geology* 13, 558-561 (1985).
11. Cochran, K. A. & Elmore, R. D. *J. sedim. Petrol.* (in the press).
12. Elmore, R. D., Cochran, K. A. & Nick, K. E. *Eos* 67, 265 (1986).
13. Zijderveld, J. D. A. in *Methods in Paleomagnetism* (eds Collinson, D. W., Creer, K. M. & Runcorn, S. K.) 254-286 (Elsevier, New York, 1967).
14. McCabe, C. *Eos* 67, 265 (1986).
15. Allan, J. R. & Mathews, R. K. *Geology* 5, 16-20 (1977).

A land-bridge island perspective on mammalian extinctions in western North American parks

William D. Newmark

School of Natural Resources, University of Michigan, Ann Arbor, Michigan 48109, USA

In recent years, a number of authors¹⁻³ have suggested several geometric principles for the design of nature reserves based upon the hypothesis that nature reserves are analogous to land-bridge islands. Land-bridge islands are islands that were formerly connected to the mainland and were created by a rise in the level of the ocean. Land-bridge islands are considered supersaturated with species in that the ratio of island to mainland species numbers is higher than expected from the area of the island. As a result, the rate of extinction should exceed the rate of colonization on a land-bridge island, resulting in a loss of species that is suggested to be related to the size and degree of isolation of the island⁴. If nature reserves are considered to be similar to land-bridge islands, because most are slowly becoming isolated from their surroundings by habitat disturbance outside the reserves⁴⁻⁸, several predictions follow. First, the total number of extinctions should exceed the total number of colonizations within a reserve; second, the number of extinctions should be inversely related to reserve size; and third, the number of extinctions should be directly related to reserve age. I report here that the natural post-establishment loss of mammalian species in 14 western North American national parks is consistent with these predictions of the land-bridge island hypothesis and that all but the largest western North American national parks are too small to retain an intact mammalian fauna.

I tested the land-bridge island predictions by examining the change in mammalian species number in 14 western North American national parks and park assemblages (Table 1) located

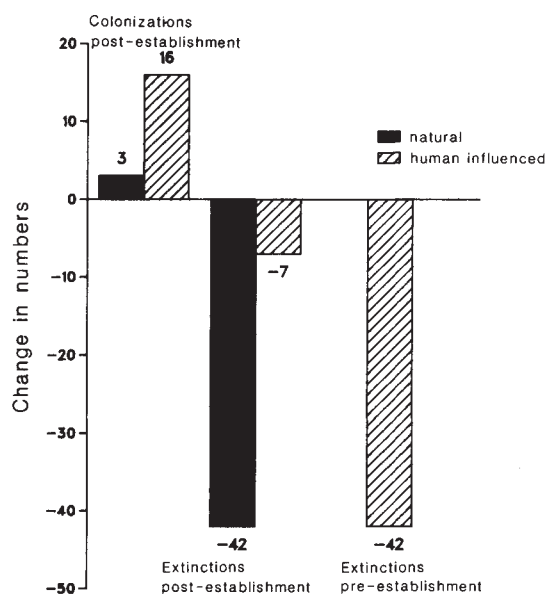


Fig. 1 Comparison of number of post-establishment colonizations with number of post-establishment and pre-establishment extinctions in 14 western North American national parks. Colonizations and extinctions are classified as either natural (solid bars) or human-influenced (hatched bars).

within the Rocky Mountains, Sierra-Cascades and Colorado Plateau. A park assemblage is defined as two or more contiguous parks. The age of a park is defined as time since park establishment; the age of a park assemblage is defined as the mean time since park establishment for the individual parks. Analysis was limited to the orders Lagomorpha, Carnivora and Artiodactyla, because these orders had the most complete park sighting records. Species of these orders tend to be more frequently reported because of their relatively large body size, non-fossorial nature and 'popularity'.

I recorded for every species in each park, based upon park sighting records and the literature, the date of last sighting as of 1983/84 and the total number of sightings for the species if it had been sighted less than five times since establishment of the park. Species which had been sighted less than three times since the establishment of the park were excluded in the analysis of post-establishment colonizations and extinctions. Biases may exist in the sighting records because of their non-standardized nature, potential misidentification of species by observers, and a lack of equivalent sampling effort between parks. I have attempted to minimize these biases by using conservative methods (see below) for classifying a sighting as valid and determining the number of post-establishment extinctions.

With the exception of sightings made by biologists or park employees, a sighting was considered valid only if it was accompanied by an accurate description of the species. The number of post-establishment extinctions was determined by assuming that all species that had not been sighted for a minimum of 10 years by 1983 were extinct. In a few cases, there was solid evidence that several species had become extinct in a park since 1973. This method for determining the number of post-establishment extinctions excludes any species that have become extinct and subsequently recolonized the park naturally since park establishment. The number of post-establishment colonizations was determined by assuming a colonization had occurred if a species had not been reported within a park near time of establishment but was subsequently sighted at least three times since park establishment. The number of pre-establishment extinctions, considered hypothetical, was determined using three criteria. First, a species must have a historic range, as described in the literature, overlapping a park. Second, the park must

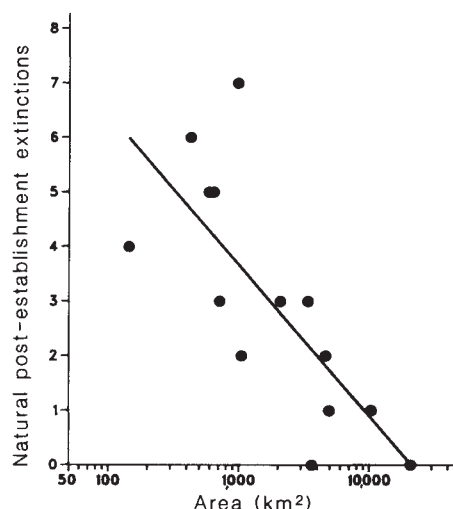


Fig. 2 Relationship between number of natural post-establishment extinctions and park area in 14 western North American national parks. Area is plotted on a logarithmic scale. The straight line shows the relationship $y = 11.95 - 2.76 \times (\log \text{area})$.

currently meet known habitat requirements for a species. Third, a species must be documented as occurring either within a park or its vicinity before park establishment by a specimen, skeletal remains, pre-establishment sighting(s), post-establishment sighting(s) fewer than three times, or a post-establishment reintroduction by a park manager.

Colonizations and extinctions were classified as being natural or human-influenced. A natural extinction or colonization was an extinction or colonization that could not be related directly to human disturbance within a park. Conversely, a human-influenced extinction or colonization was one that could be potentially or directly related to human activities within a park (see below). Data are summarized in Table 1.

The number of cases (Fig. 1) of natural post-establishment extinctions of mammals ($n = 42$) has exceeded the number of post-establishment colonizations ($n = 3$). This result is consistent with the first prediction of the land-bridge island hypothesis. The natural colonizations resulted from range expansions by the raccoon (*Procyon lotor*) and the moose (*Alces alces*). Park managers have reintroduced 12 species that were found historically in the parks and four exotic species have colonized the parks from lands adjacent to the parks. I attribute the post-establishment human-influenced extinction of six species to predator control and one species to accidental poisoning. A large number of species ($n = 42$) were also lost before the establishment of most western North American national parks (Fig. 1). However, it is quite likely that many of the species that became extinct in the smaller parks before park establishment were transient populations. The pre-establishment loss of species is most probably attributable to human disturbance in the form of hunting, logging, grazing and mining which occurred in many of the western North American parks before their establishment⁹. Combining both pre- and post-establishment extinctions, up to 43% of all species of lagomorph, carnivore and artiodactyl (12 species) found historically within 14 western North American national parks have been lost within a given park (Table 1). Only the largest western North American park assemblage, the Kootenay-Banff-Jasper-Yoho park assemblage (20,736 km²), still contains an intact historical mammalian faunal assemblage.

The number of natural post-establishment mammalian extinctions (Fig. 2) is significantly and inversely related to log park area ($r^2 = 0.56$, $P < 0.01$). The number of natural post-establishment mammalian extinctions is insignificantly correlated with log park age ($r^2 = 0.00$, $P > 0.97$). However, if the effects of park

Table 1 Summary of pre- and post-establishment extinctions and colonizations, area and age for 14 western North American national parks

Park/Park assemblage	Proportional loss of total species found historically (%)	No. of pre-establishment human-influenced extinctions	No. of post-establishment natural extinctions	No. of post-establishment human-influenced extinctions	No. of post-establishment natural colonizations	No. of post-establishment human-influenced colonizations	Area (km ²)	Age (years)
Bryce Canyon	36	4	4	1	0	1	144	61
Lassen Volcanic	43	6	6	0	0	0	426	77
Zion	36	5	5	0	0	2	588	75
Crater Lake	31	5	5	0	0	2	641	82
Manning Provincial	26	4	3	0	1	1	712	43
Mount Rainier	32	1	7	0	0	1	976	85
Rocky Mountain	31	10	2	0	0	3	1,049	69
Yosemite	25	2	3	1	0	0	2,083	94
Sequoia-Kings Canyon	23	1	3	3	0	1	3,389	94
Olympic	6	0	0	1	0	2	3,628	75
Glacier-Waterton Lakes	7	1	2	0	1	1	4,627	81.5
Grand Canyon	18	2	1	1	0	0	4,931	76
Grand Teton-Yellowstone	4	0	1	0	1	1	10,328	83.5
Kootenay-Banff-Jasper-Yoho	0	1	0	0	0	1	20,736	84.5

Data from ref. 13.

area are held constant, the number of natural post-establishment extinctions is significantly and positively partially correlated with log park age ($r = 0.57$, $P < 0.05$). Log park area (a) and log park age (g) when combined in a multiple linear regression model account for 71% of the total variation in numbers of natural post-establishment extinctions (e). The multiple regression model is: $e = -6.04 - 3.49(a) + 10.82(g)$ ($n = 14$, $F = 13.15$, $P < 0.002$). The results of the simple linear regression and the partial correlation are consistent with the second and third predictions of the land-bridge island hypothesis. In addition, the results of the simple and multiple linear regression indicate that park area is a more important determinant than park age of number of natural post-establishment mammalian extinctions in western North American parks.

The natural post-establishment loss of mammalian species is most probably attributable to the loss of habitat and the active elimination of fauna on adjacent lands or what has been described as short-term insularization effects¹⁰. This loss of habitat and the active elimination of fauna on lands adjacent to the parks have had a twofold effect. First, they have increased the probability of local extinction of species within the reserves because smaller parks tend to have smaller populations which in turn have a higher probability of extinction^{11,12}. Further support for this hypothesis is provided by the fact that population size is the most consistent predictor in a multivariate statistical analysis of life history characteristics of the probability of post-establishment persistence for populations of lagomorphs, carnivores and artiodactyls within 24 western North American national parks¹³. Second, this disturbance has reduced the potential for colonization from lands adjacent to the parks by increasing the distance or isolation of the parks from potential source areas. Openings as narrow as a road, open field, or clearcut have been shown to inhibit the movement of both large and small mammals¹⁴⁻¹⁷.

Without active intervention by park managers, it is quite likely that a loss of mammalian species will continue as western North American national parks become increasingly insularized. Yet it is possible that even without further insularization, mammalian species may continue to be lost in these parks because of a lag between past disturbance and subsequent extinction. Augmentation of existing mammalian park populations by natural colonization, or what has been described as the rescue effect¹⁸, probably will be limited to the most common species. The enhancement of populations for rare species will be largely dependent upon the active introduction of individuals by park managers.

A final factor that could explain the natural post-establishment loss of mammalian species in western North American parks is habitat change. It is widely acknowledged that fire suppression since the early 1900s has affected the vegetative structure of

western North American national parks¹⁹⁻²². However if this vegetative change has had an influence on post-establishment mammalian extinctions, one would expect that species that are dependent upon early successional vegetation should be most prone to extinction. Of the 42 populations of species that have become extinct in 14 parks since park establishment, only 10 may be classified as being dependent upon early successional vegetation. Thus for most species, it appears that successional change is not the principal determinant of post-establishment local extinction of species.

The natural post-establishment loss of mammalian species in western North American national parks indicates that virtually all western North American national parks were too small to maintain the mammalian faunal assemblage found at time of park establishment. To reduce the potential loss of mammalian fauna in the future will most probably require that the mammalian fauna within the parks be more actively managed and that the parks be 'enlarged' either through the acquisition or the cooperative management of lands adjacent to the parks.

I thank G. E. Belovsky, K. R. Miller, M. E. Soulé, B. J. Rathcke and B. S. Low for their advice and H. L. Gibbs for comments. I also thank the many National Park Service and Environment Canada, Parks employees who graciously reviewed species checklists and provided requested information. Support was provided by two Rackham Dissertation Improvement grants and four School of Natural Resources Thesis grants.

Received 23 July; accepted 24 November 1986.

1. Diamond, J. M. *Biol. Conserv.* 7, 129-146 (1975).
2. Terborgh, J. in *Tropical Ecological Systems: Trends in Terrestrial and Aquatic Research* (eds Golley, F. B. & Medina, E.) 369-380 (Springer, New York, 1975).
3. Wilson, E. O. & Willis, E. O. in *Ecology and Evolution of Communities* (eds Cody, M. L. & Diamond, J. M.) 522-534 (Harvard University Press, Cambridge, 1975).
4. Diamond, J. M. *Proc. natn. Acad. Sci. U.S.A.* 69, 3199-3203 (1972).
5. Wright, G. M., Dixon, J. S. & Thompson, B. H. *Fauna Series No. 1, Vol. 1* (US Govt Printing Office, Washington DC, 1932).
6. Soulé, M. E., Wilcox, B. A. & Holtby, C. *Biol. Conserv.* 15, 259-272 (1979).
7. National Park Service, *State of the Parks - 1980; A Report to the Congress* (US Dept of the Interior, Washington DC, 1980).
8. Newmark, W. D. *Biol. Conserv.* 33, 197-208 (1985).
9. Runte, A. *National Parks: The American Experience* (University of Nebraska, Lincoln, 1979).
10. Wilcox, B. A. in *Conservation Biology: An Evolutionary-Ecological Perspective* (eds Soulé, M. E. & Wilcox, B. A.) 95-117 (Sinauer, Sunderland, 1980).
11. Terborgh, J. & Winter, B. in *Conservation Biology: An Evolutionary-Ecological Perspective* (eds Soulé, M. E. & Wilcox, B. A.) 119-133 (Sinauer, Sunderland, 1980).
12. Diamond, J. M. in *Extinctions* (ed. Nitecki, M. H.) 191-246 (University of Chicago Press, Chicago, 1984).
13. Newmark, W. D. Ph.D., thesis, Univ. Michigan (1986).
14. Mader, H. J. *Biol. Conserv.* 29, 81-96 (1984).
15. Oxley, D. J., Fenton, M. D. & Carmondy, G. R. *J. appl. Ecol.* 11, 51-59 (1974).
16. Pedersen, R. J. in *Conf. Proc. Recreational Impact of Wildlands* (eds Ittner, R., Potter, D. R., Agee, J. K. & Anshell, S.) 169-173 (US Forest Service, 1979).
17. Wegner, J. F. & Merriam, G. *J. appl. Ecol.* 16, 349-357 (1979).
18. Brown, J. H. & Kodrick-Brown, A. *Ecology* 58, 445-449 (1977).
19. Kilgore, B. M. & Taylor, D. *Ecology* 60, 129-142 (1979).
20. Loope, L. L. & Gruell, G. E. *Quat. Res.* 3, 425-443 (1973).
21. McNeil, R. C. & Zobel, D. B. *NW Sci.* 54, 30-46 (1980).
22. Vankat, J. L. & Major, J. *J. Biogeogr.* 5, 377-402 (1978).

TIT FOR TAT in sticklebacks and the evolution of cooperation

Manfred Milinski

Arbeitsgruppe für Verhaltensforschung, Fakultät für Biologie,
Ruhr-Universität, Postfach 102148, 4630 Bochum 1, FRG

The problems of achieving mutual cooperation can be formalized in a game called the Prisoner's Dilemma in which selfish defection is always more rewarding than cooperation¹. If the two protagonists have a certain minimum probability of meeting again a strategy called TIT FOR TAT is very successful². In TIT FOR TAT the player cooperates on the first move and thereafter does whatever the opponent did on the previous move. I have studied the behaviour of fish when confronting a potential predator, because conflicts can arise within pairs of fish in these circumstances which I argue resemble a series of games of Prisoner's Dilemma. Using a system of mirrors, single three-spined sticklebacks (*Gasterosteus aculeatus*) approaching a live predator were provided with either a simulated cooperating companion or a simulated defecting one. In both cases the test fish behaved according to TIT FOR TAT supporting the hypothesis that cooperation can evolve among egoists.

Cooperation between individuals which are not closely related has been a difficulty for evolutionary theory until recently. Following Trivers's theory of reciprocal altruism³, Axelrod and Hamilton² used a game theoretical approach including computerized tournaments of iterated games of Prisoner's Dilemma. In this game, two individuals have to choose either to cooperate or to defect. If they cooperate, both do better than if both had defected. But if one player defects while the other cooperates, the defecting one gets more than if both had cooperated, and the cooperating one gets less than if both had defected. Therefore, if the reward for mutual cooperation is R units, for mutual defection is P units, and for cooperation-defection is T units to the defector and S to the cooperator, then $T > R > P > S$ and $R > (S + T)/2$ are the conditions for the dilemma to exist. It pays always to defect if there is only one encounter or a previously known number of encounters between two particular individuals. If there is a certain minimum probability that after the current interaction the two protagonists will meet again, a very simple strategy, called TIT FOR TAT is demonstrably superior²: cooperate on the first move, on all subsequent moves do what the other player did on the preceding one. TIT FOR TAT players were successful because they were never the first to defect, they immediately retaliated when provoked, and were forgiving after just one act of retaliation².

A variety of situations should give rise to TIT FOR TAT-like strategies for cooperation, but so far the only experimental support is from tree swallows (*Tachycineta bicolor*)⁴. A further requirement must be satisfied before cooperation can evolve^{1,5}: the gains expected from future encounters between any two protagonists must not be too heavily discounted. In many biological applications the future is more or less uncertain, so that future gains may be worth less than current ones. This paper investigates a situation in which future gains are worth more than current ones, enhancing the likelihood that TIT FOR TAT will occur.

During the early stages of an attack by a stalking pike (*Esox lucius*), individual minnows (*Phoxinus phoxinus*) leave the shoal and approach to within 4–6 body lengths of the predator, wait there for a few seconds, then slowly turn and go back to the shoal^{6–8}. This behaviour, reported first by George⁹, has been termed a predator inspection visit¹⁰ and it has been suggested that fish gather knowledge about the predator's identity, precise location and current motivational state¹¹ from such visits. Of course there is an increasing danger of being attacked with

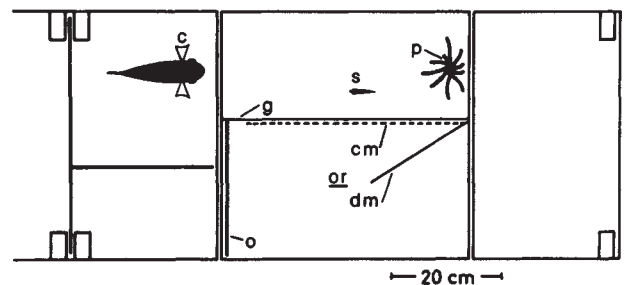


Fig. 1 Plan of the experimental equipment. The glass tanks shown contain the cichlid (c) and the plant (p), respectively. A typical position of a test stickleback (s) is also shown, as are the positions of the cooperating mirror (cm) and the defecting mirror (dm); 0, opaque screen; g, glass partition. Only one of the two mirrors was present in each experiment.

decreasing distance to the predator, although the payoff (the amount of information gathered) may be maximal close to the predator. Similar behaviour has also been described in three-spined sticklebacks¹².

Three-spined sticklebacks approach a live cichlid more closely when in a group of two than when alone¹³, like minnows approaching a pike model¹⁰ and paired birds attacking or mobbing a predator^{14,15}. The two sticklebacks remain close to each other and move in a jerky way: short moves of a few centimetres towards the predator alternate with hesitations. Either the two fish alternated positions or, more frequently, one followed the other. Although both fish may benefit from selfish herd effects¹⁶ or predator confusion¹¹, one or other fish has to be first to make a move forward running a higher risk of being preyed upon. Both fish finally rush back when the predator moves towards them. If each forward jerk of swimming is one episode of cooperation, then the entire inspection visit of the two fish is equivalent to a sequence of encounters. The payoff (information gathered) from each episode should increase with decreasing distance to the predator. For the fish which follows, staying behind (defecting) may decrease its risk but increases its payoff by watching the fate of the other fish. This investigation tests whether the leading fish adopts TIT FOR TAT when a simulated companion either follows immediately behind it (cooperates) or stays behind and ultimately disappears (defects).

The experimental tank (43 × 43 cm, water level 18 cm) was divided by a glass partition into two compartments: the experimental compartment containing a plant (*Vallisneria* sp.) at one end and another compartment containing a mirror (Fig. 1). In one experiment there was a long mirror (38 cm) just behind the glass partition and parallel to it. This 'cooperating mirror' simulated a companion which followed immediately a proceeding stickleback. In another experiment there was a short mirror (19 cm) at an angle of 32° to the glass partition, called the 'defecting mirror' because it simulated a companion which stayed increasingly behind a proceeding stickleback. The experimental tank was positioned between two other tanks one of which contained a tame cichlid (*Tilapia mariae*, ~18 cm long) which resembles a perch (*Perca fluviatilis*), a common predator of three-spined sticklebacks¹⁷. Light was provided by four fluorescent tubes mounted 3 m above the tanks. The outside walls were covered with grey paper to avoid visual distraction.

Before a trial, a stickleback (the first to swim voluntarily into the net in the storage tank) was released above the plant in the experimental compartment. Trials of the two types of experiment were alternated and the length distribution of the fish was similar in both experiments (4.0 ± 0.6 cm, mean $\pm$ s.d., $n = 25$). The experiments were made outside the breeding season and each stickleback was used once only. Both the cichlid and the stickleback were observed with a video camera suspended 1.5 m above the tanks. On the video screen, the length of the experimental compartment was subdivided into 20 sections of 2.15 cm each

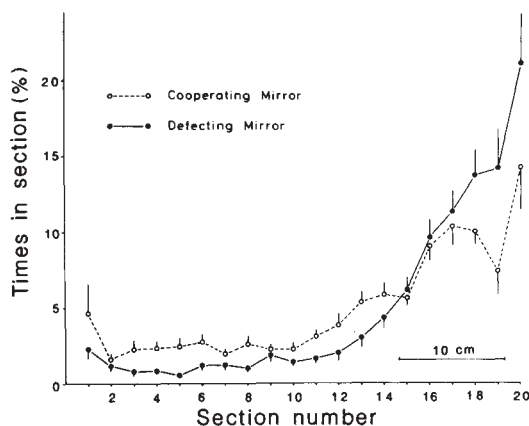


Fig. 2 Mean percentage of occasions on which fish with (○, broken line) the cooperating mirror ($n=25$) and fish with (●, solid line) the defecting mirror ($n=25$) were found in each of the 20 sections representing different distances to the predator; bars, 1 s.e.m.

with section 1 nearest the cichlid. The section containing the tip of the stickleback's snout was recorded at a clock signal every 5 s during 5 min after the fish had been transferred to the experimental compartment. The results therefore showed on how many occasions out of a total of 60 the fish had been found in each section.

The cichlid spent most of its time close to the wall where it watched the stickleback, and often pushed the wall as it tried to attack. The stickleback usually stayed close to the plant after being transferred, but after some hesitation approached the cichlid in its typical jerky manner along the glass partition. Normally it stopped after a distance of a few centimetres and retreated abruptly, especially when the cichlid attacked it. Some fish approached no more closely than to section 15, whereas others reached section 1 and stayed there for >5 s, rushing back as soon as they were attacked.

Although in both experiments the sticklebacks preferred to stay in the back half of their tank, there were differences between experiments. With the cooperating mirror the sticklebacks were twice as often in the front half (mean 25.1%) as with the defecting mirror (mean 12.1%, $P<0.003$, two-tailed Mann-Whitney U -test). Figure 2 shows how often fish in each experiment were found in each of the 20 sections. With the cooperating mirror the fish were found significantly more frequently in each of the 3 front quarters than in the experiment with the defecting mirror (section 1-5, $P<0.05$; section 6-10, $P<0.004$; section 11-15, $P=0.04$; two-tailed Mann-Whitney U -tests). With the defecting mirror, fish were found more often in the back quarter of the tank than in the experiment with the cooperating mirror ($P<0.005$, two-tailed Mann-Whitney U -test).

In both experiments the sticklebacks decreased their mean distance to the cichlid significantly during the 5 min of the trial (Fig. 3). The fish with the cooperating mirror had not only a shorter mean distance to the predator in each half minute of the trial but also decreased their distance more than proportionally from the first to the fifth minute as compared to the fish with the defecting mirror (the ratio of mean section during the first minute to the mean section during fifth minute was 1.54 with the cooperating mirror and 1.17 with the defecting mirror, $P<0.05$, one-tailed Mann-Whitney U -test). There are consistent differences between individual fish. In the experiment with the cooperating mirror there is a significant positive correlation between distance in the first minute and in the fifth minute ($r=0.43$ with 23 d.f., $P<0.04$, two-tailed test). Thus, there were bold fish and more cautious ones. This correlation is not significant for fish with the defecting mirror ($r=0.34$ with 23 d.f., $P>0.07$, two-tailed). The 13 fish which had a shorter mean

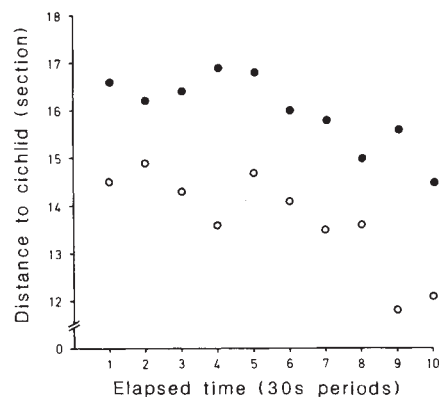


Fig. 3 Mean distance to cichlid (section number) of fish with cooperating mirror and of fish with defecting mirror in each half min of the trial; for the defecting mirror, $r_s=-0.79$, $P<0.01$; for cooperating mirror, $r_s=-0.83$, $P<0.005$ (two-tailed test with 23 d.f.).

distance during the first minute are called 'bold' and the remaining 12 fish 'cautious' in either experiment. With the cooperating mirror both bold and cautious fish significantly decreased their mean distance from the first to the fifth minute (Fig. 4), although bold fish were still closer to the cichlid than cautious ones by the fifth minute ($P<0.04$, one-tailed Mann-Whitney U -test). In the experiment with the defecting mirror, however, only the cautious fish significantly decreased their mean distance to the predator (Fig. 4): bold fish were not significantly closer to the cichlid than the cautious ones by the fifth minute ($P>0.1$, one-tailed Mann-Whitney U -test).

Three-spined sticklebacks are able to adjust their behaviour according to the risk of predation¹⁸. In both experiments the safest place for the stickleback was close to the back wall of the experimental compartment. This was also the place where the companion fish, simulated by the mirror, was very close. Nevertheless, in both experiments the test sticklebacks left this place to make inspection visits towards the cichlid. In the experiment with the long parallel mirror, a cooperating companion was simulated which followed the proceeding stickleback immediately, that is, it adopted TIT FOR TAT. In the experiment with the short oblique mirror, a companion was simulated which stayed increasingly behind the stickleback the further the fish advanced, and it disappeared half-way to the cichlid, so that it

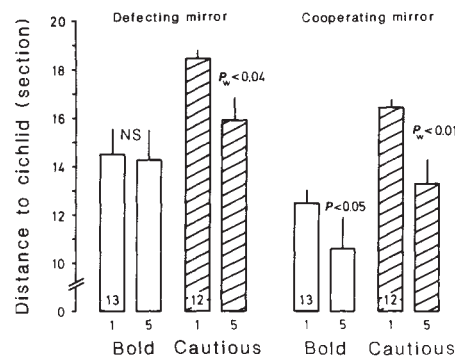


Fig. 4 Mean distance to cichlid (section number) of (open bars) bold and of (hatched bars) cautious fish in the first and fifth minute of the trial in the experiment with the defecting mirror and in the experiment with the cooperating mirror; bars above columns give 1 s.e.m. Wilcoxon matched pairs signed-ranks test (two-tailed) gives for cautious fish, $P_w<0.04$ with defecting mirror; $P_w<0.01$ with cooperating mirror. Dixon and Mood sign-test (two-tailed) gives for bold fish, no significant difference between minutes 1 and 5 with the defecting mirror; $P<0.05$ with the cooperating mirror.

mostly defected. If the sticklebacks adopt TIT FOR TAT in this situation, with each jerky approach of some centimetres towards the cichlid representing one opportunity for cooperation, then fish with the cooperating mirror should approach the predator more closely than those with the defecting mirror. This was the case. The sticklebacks acted as if they perceived that a companion was either following them or staying increasingly behind.

Individual sticklebacks differ consistently in their levels of boldness^{12,19} which seems generally typical for fish behaviour^{20,21}. With the cooperating mirror, the sticklebacks that were bolder in the first minute were also bolder in the fifth minute of the trial, although the cautious fish also approached the predator more and more closely during the trial. Hence, despite individual differences in boldness, all the fish seemed to adopt the same strategy when the simulated companion cooperated. This was different when a defecting companion was simulated. Here only the cautious fish approached the cichlid a bit more closely in the fifth than in the first minute. Over this distance the simulated companion did not stay much behind. The bolder fish, however, reacted to the disappearance of their companion in that they neither decreased nor increased their mean distance from the cichlid during the trial. Thus, they seem to 'forgive' after once having rushed back themselves. They went forward time and again, but on the average exactly as far in the fifth minute as they had done in the first minute. So a further requirement for TIT FOR TAT seems to be fulfilled, the equivalent of cooperation on the first move.

Whether the condition $T > R > P > S$ is fulfilled was not determined quantitatively; there is, however, qualitative support for it. The bold fish with the cooperating mirror seemed to rush back mainly when the cichlid actually launched an attack. Therefore, the ultimate information the fish try to gather seems to be the predator's attack distance. The rearmost fish achieves this by watching from a safe distance what happens to the leading fish, so $T > R$ seems to be fulfilled. $R > P$ seems to hold because if both fish stay away, they do not learn about the predator's attack distance. $P > S$ holds because if the leading fish relies on confusion or the selfish herd effect it may approach close enough to learn the predator's attack distance but runs a high risk of being eaten. $R > (S + T)/2$ should also be fulfilled. A group of two fish may detect an impending attack earlier⁶ which may be less frequently successful anyway because of the confusion effect of two prey compared to one²². Either cooperating fish may gather maximum information while running less than half of the predation risk of a single fish. Thus, the sticklebacks support Axelrod and Hamilton's assumptions and predictions² concerning the evolution of cooperation based on reciprocity.

I thank R. Axelrod, W. D. Hamilton, A. E. Magurran and T. J. Pitcher for comments on the manuscript, and C. Wynne for improving my English.

Received 15 September; accepted 21 November 1986.

- May, R. M. *Nature* **292**, 291-292 (1981).
- Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390-1396 (1981).
- Trivers, R. Q. *Rev. Biol.* **46**, 35-57 (1971).
- Lombardo, M. P. *Science* **227**, 1363-1365 (1985).
- Axelrod, R. *Am. pol. Sci. Rev.* **75**, 306-318 (1981).
- Magurran, A. E., Oulton, W. J. & Pitcher, T. J. *Z. Tierpsychol.* **67**, 167-178 (1985).
- Magurran, A. E. & Girling, S. L. *Anim. Behav.* **34**, 510-518 (1986).
- Magurran, A. E. & Pitcher, T. J. *Proc. R. Soc. (in the press)*.
- George, C. J. W. Thesis, Harvard Univ. (1961).
- Pitcher, T. J., Green, D. A. & Magurran, A. E. *J. Fish Biol.* **28**, 439-448 (1986).
- Pitcher, T. J. in *The Behaviour of Teleost Fishes* (ed. Pitcher, T. J.) 294-337 (Croom Helm, London, 1986).
- Huntingford, F. A. *Anim. Behav.* **24**, 245-260 (1976).
- Milinski, M. *Behaviour* **93**, 203-216 (1985).
- Bossemma, I. & Benus, R. F. *Behav. ecol. Sociobiol.* **16**, 99-104 (1985).
- Regelmann, K. & Curio, E. *Behaviour* **97**, 10-36 (1986).
- Hamilton, W. D. *J. theor. Biol.* **31**, 295-311 (1971).
- Wotton, R. J. *The Biology of the Sticklebacks* (Academic, London, 1976).
- Milinski, M. & Heller, R. *Nature* **275**, 642-644 (1978).
- Huntingford, F. A. *Anim. Behav.* **30**, 909-916 (1982).
- Magurran, A. E. *Behav. ecol. Sociobiol.* **19**, 267-273 (1986).
- Magurran, A. E. in *The Behaviour of Teleost Fishes* (ed. Pitcher, T. J.) 338-365 (Croom Helm, London, 1986).
- Ohguchi, O. *Prey Density and Selection against Oddity by Three-spined Sticklebacks* (Parey, Berlin, 1981).

Efficiency characteristics of crescent-shaped wings and caudal fins

C. P. van Dam

Department of Mechanical Engineering, Division of Aeronautical Science and Engineering, University of California Davis, Davis, California 95616, USA

Caudal (tail) fins of fish and aquatic mammals that cruise long distances, and wings of certain birds, often have the shape of a crescent moon. This study investigates how the crescent shape contributes to the travelling performance of these animals. A steady-flow theory¹ that correctly models the trailing wake was used to analyse lifting surface efficiency, which is dependent on the level of induced (or vortex) drag for a given lift and span of the lifting surface. This analysis shows that backward curvature of a wing improves induced efficiency to a value greater than that of the flat untwisted wing of elliptical shape considered optimal in classical wing theory^{2,3}. This increase of induced efficiency results from the nonplanar trailing vortex sheet produced by the crescent-shaped wing at a given angle of attack.

Many aquatic animals that swim fast for long distances have adopted the mode of propulsion Lighthill⁴ terms carangiform with lunate (crescent) tail. Forward thrust is generated exclusively by the caudal fin mounted on the slender tail section of the heavy body. Scombroid fishes (Fig. 1a) have the caudal fin strengthened by ossified fin rays allowing this fin to have a high aspect ratio⁵. Fast-swimming sharks (Fig. 1b) and cetaceans (Fig. 1c) have lunate tails of a lower aspect ratio because of lower structural strength and stiffness in their caudal fins.

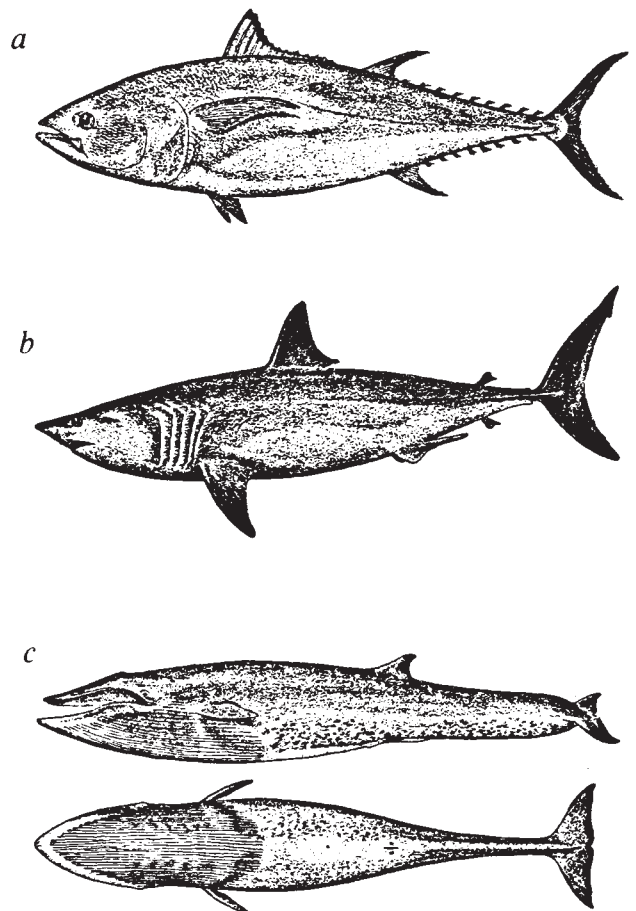


Fig. 1 Typical aquatic animals with crescent-shaped caudal fins¹⁹: a, *Thunnus thynnus*; b, *Isurus oxyrinchus*; c, *Balaenoptera borealis*.

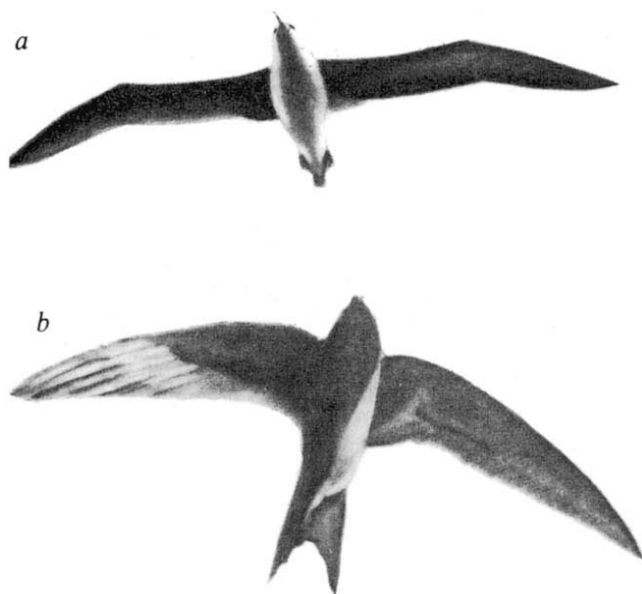


Fig. 2 Birds with backward wing curvature: a, *Diomedea immutabilis*; b, *Apus apus*.

Birds that have encountered few environmental restrictions limiting wing shape have also evolved the crescent shape with high aspect ratio that gives improved aerodynamic performance according to the present study. Many sea birds have relatively long, heavy wings with highly tapered, aft-swept tips (Fig. 2a). Some land birds such as the swift (Fig. 2b) also have wings of this shape⁶. Swifts must be aerodynamically very efficient because, feeding on the wing, they spend most of their lives in flight.

Classical steady wing theory does not predict improved aerodynamic efficiency for lifting surfaces of crescent shape when compared with the optimal straight wing of elliptical shape^{7,8}. Unsteady two-dimensional theory is inadequate to predict the efficiency of lunate tails (or crescent wings)^{4,9-11} because it completely neglects the influence of the trailing vortex sheet. Unsteady three-dimensional methods¹²⁻¹⁶ also fail to predict adequately the increased efficiency of the lunate shape in oscillatory motion at the frequencies observed in these animals. These failures to predict the superior efficiency characteristics of crescent-shaped lifting surfaces may be caused by the inability of all these computational models to account properly for the trailing wake.

The present study will show that crescent planform shapes can significantly increase the level of induced efficiency for lifting surfaces even in steady flow. Results published by Zimmer¹⁷, although calculated differently, indicate similar trends.

The steady-flow analysis is conducted for a constant angle of attack. The reduced frequency parameter $\nu = \omega c_{av} / U_\infty$ for oscillating caudal fins and flapping wings is generally less than 0.6, where $\omega (=2\pi f)$ is the radian frequency, c_{av} is the mean chord, and U_∞ represent the freestream velocity. Therefore, for attached-flow conditions the unsteady effects are relatively small and a quasi-steady analysis is adequate to investigate the efficiency of crescent shapes⁶.

The induced drag of a wing can be calculated as follows:

$$D_i = 2L^2 / (\pi \rho U_\infty^2 b^2 e)$$

where L denotes the lift produced by the wing and ρ the fluid density. Induced drag is thus inversely proportional to the square of the geometric wing span b and also inversely proportional to the induced efficiency e . Within the framework of Prandtl's

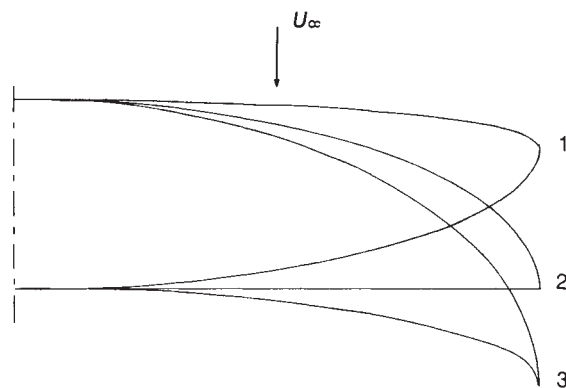


Fig. 3 Top views of three wing configurations with identical spanwise chord distributions and differing backward curvatures. Wing 1 has the classical elliptical configuration with straight quarter-chord line, wing 2 has straight trailing edge and wing 3 has a typical crescent shape. Each wing aspect ratio ($=b^2/S$) is 7 with symmetrical NACA0012 (ref. 20) section shape in the streamwise direction.

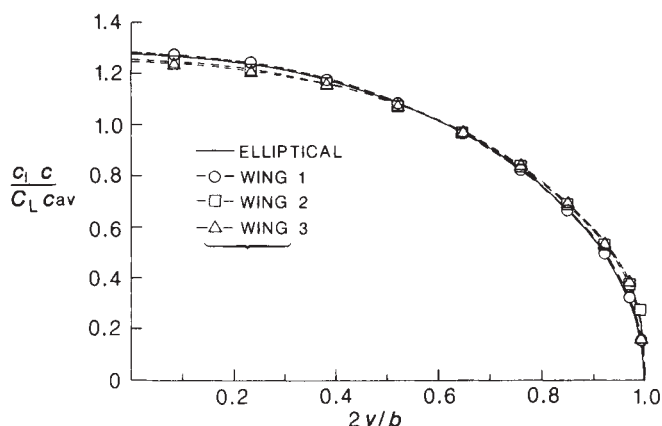


Fig. 4 Spanwise load distributions of wings at $\alpha = 4.0^\circ$, where c_l denotes the section lift coefficient at spanwise location y and c the chord length at spanwise location y . An elliptical load distribution is given for comparison.

classical lifting-line theory², the maximum value for e of 1.0 is obtained for a straight-planar wing with an elliptical planform. The wake of the wing is modelled as a flat trailing vortex sheet and this wake model is also often applied in lifting-surface theory^{14,15}. In the tip regions, however, vortex roll-up occurs along the free edges of the trailing wake and this influences the aerodynamic loading of the wing. More importantly, at angle of attack the trailing edge of a crescent-shaped wing and the wake which springs from the trailing edge are highly nonplanar. This geometric feature is not well modelled by methods that assume the trailing wake to be flat.

The surface panel method¹ correctly models the trailing wake and its influence on the wing-loading distribution (see for example ref. 18). A very important feature for this study is that the method allows for wake relaxation and vortex roll-up. A rapidly converging iterative procedure aligns the wake panels with the local flow condition and provides for the essential nonlinear effect of the nonplanar trailing wake. Typically, 3-5 iterations are required to obtain the correct wake shape and, thus, a converged solution.

First, the classical optimal wing with its elliptical planform shape and unswept quarter-chord line is analysed (wing 1 in Fig. 3). A total of 1,000 panels are generated to model the lifting surface assuming geometric and aerodynamic symmetry about the centreline. At an angle of attack $\alpha = 4^\circ$, the potential-flow

method gives a lift coefficient $C_L = 0.34175$ and an induced-drag coefficient $C_{Di} = 0.00531$, where $C_L = 2L/(\rho U_\infty^2 S)$ and $C_{Di} = 2D_i/(\rho U_\infty^2 S)$, and S is the planform area. When these values are substituted in the induced-drag formula, an induced efficiency $e = 1.0$ is obtained, as expected. The spanwise load distribution for wing 1 (Fig. 4) shows good agreement with the elliptical load distribution plotted.

Next, two wings with different amounts of sweep are analysed to examine the effect of backward curvature on their induced-drag characteristics when compared with a wing of 'optimal' planform (Fig. 3). The two aft-swept wing shapes are obtained from the original wing shape by 'shearing' it so that the chord lengths parallel to the centreline, the wing span and the planform area are all unchanged. In addition, panel number and distribution are held constant. Wing 2 has a backward-curved leading edge and straight trailing edge. At constant $\alpha = 4^\circ$, $C_L = 0.34236$ and $C_{Di} = 0.00510$ resulting in $e = 1.045$; a 4.5% increase in induced efficiency compared to wing 1. For wing 2 the spanwise load distribution is no longer elliptical (Fig. 4). For a constant chord distribution, backward sweep produces an increase in aerodynamic loading outboard towards the tip (decrease inboard towards the root) of the lifting surface. Wing 3 has both leading edge and trailing edge curved backwards. The change in trailing edge from straight to curved backwards has relatively little influence on the load distribution (Fig. 4), but causes a slight reduction in lift-curve slope and a significant reduction in induced drag. At constant $\alpha = 4^\circ$, $C_L = 0.33714$ and $C_{Di} = 0.00475$ resulting in $e = 1.088$; an 8.8% increase in induced efficiency over wing 1. The spanwise chord distribution and the resulting load distribution are not considered to be optimal. The

problem of determining the optimal load distribution for crescent-shaped lifting surfaces is still being studied.

In conclusion, this steady-flow analysis of lift and drag characteristics of crescent-shaped wings indicates that aerodynamic efficiency is improved as a result of increased backward curvature. The combination of high induced efficiency and large wing span results in low drag from lift for a given loading. This may help explain why fast marine animals and efficiently flying birds have evolved long crescent-shaped lifting surfaces.

This research was supported in part by the University of California, Davis (FRG-D-1967) and by NASA Langley Research Center.

Received 15 September; accepted 1 December 1986.

1. Maskew, B. J. *Aircraft* **19**, 157-163 (1982).
2. Prandtl, L. *NACA Rep.* **116**, (1921).
3. Munk, M. M. *NACA Rep.* **121**, (1921).
4. Lighthill, M. J. *A. Rev. Fluid Mech.* **1**, 413-446 (1969).
5. Fierstine, H. L. & Walters, V. *Mem. S. Calif. Acad. Sci.* **6**, 1-31 (1968).
6. Lighthill, M. J. in *Swimming and Flying in Nature* Vol. 2 (eds Wu, T. Y., Brokaw, C. J. & Brennen, C.) 423-491 (Plenum, New York, 1975).
7. Ashenberg, J. & Weihs, D. *Israel J. Techn.* **20**, 160-165 (1982).
8. Ashenberg, J. & Weihs, D. *J. Aircraft* **21**, 89-91 (1984).
9. Lighthill, M. J. *J. Fluid Mech.* **44**, 265-301 (1970).
10. Wu, T. Y. *J. Fluid Mech.* **46**, 337-355 (1971).
11. Wu, T. Y. *J. Fluid Mech.* **46**, 521-544 (1971).
12. Chopra, M. G. in *Swimming and Flying in Nature* Vol. 2 (eds Wu, T. Y., Brokaw, C. J. & Brennen, C.) 635-650 (Plenum, New York, 1975).
13. Chopra, M. G. *J. Fluid Mech.* **64**, 375-391 (1974).
14. Chopra, M. G. & Kambe, T. *J. Fluid Mech.* **79**, 49-69 (1977).
15. Lan, C. E. *J. Fluid Mech.* **93**, 747-765 (1979).
16. Cheng, H. K. & Murillo, L. E. *J. Fluid Mech.* **143**, 327-350 (1984).
17. Zimmer, H. thesis, Univ Stuttgart (1983).
18. Van Dam, C. P. *S.A.E. Trans.* **94**, 355-364 (1985).
19. Norman, J. R. & Fraser, F. C. *Giant Fishes, Whales and Dolphins* (Putnam, London, 1937).
20. Abbott, J. H. & Von Doenhoff, A. E. *Theory of Wing Sections* (Dover, New York, 1959).

Dopamine enhances excitatory amino acid-gated conductances in cultured retinal horizontal cells

Andrew G. Knapp & John E. Dowling

Department of Cellular and Developmental Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

In the teleost retina, cone horizontal cells receive extensive innervation from dopaminergic interplexiform cells¹, and possess dopamine receptors whose activation stimulates adenylate cyclase^{2,3}. Exogenously applied dopamine modifies several aspects of horizontal cell activity in the intact retina, including the responsiveness of these neurons to light⁴⁻⁶ and the strength of electrical coupling between them⁷⁻⁹. We have used whole-cell voltage clamp methods to examine whether dopamine can alter the light-responsiveness of horizontal cells by changing their sensitivity to the neurotransmitter released by the photoreceptors. We report that dopamine and cyclic AMP, although having little direct effect on resting membrane conductance, greatly enhance ionic conductances gated by kainate, an agonist of the transmitter released by the photoreceptors, and by L-glutamate, the agent proposed to be the photoreceptor transmitter¹⁰⁻¹⁴. Our results provide the first direct evidence for dopaminergic regulation of excitatory amino-acid neurotransmission in the vertebrate nervous system and suggest a possible mechanism to explain the reduction in the responsiveness of horizontal cells observed when retinas are treated with dopamine.

Experiments were performed on retinal horizontal cells isolated from the white perch (*Roccus americana*). These neurons retain their distinctive morphologies in short-term culture (Fig. 1, legend) allowing identification of rod-driven as opposed to cone-driven cells (ref. 15 and unpublished data). In agreement with voltage clamp⁹ and current clamp¹³ studies, we found that application of a pulse of Ringer's containing dopamine

(200 μ M) to isolated horizontal cells had no consistent effects on resting membrane conductance (Fig. 1a). In 42% of the cone horizontal cells studied ($n = 38$), dopamine caused a small outward current, whereas in 24% of cells we observed a similarly small inward current. In the remaining 34% of the cells, dopamine either produced no discernible current or biphasic responses. Similar results were obtained over a wide range of holding potentials (data not shown). Dopamine receptors on these neurons thus do not appear to be directly linked to ionic conductances. In contrast, application of a pulse of Ringer's containing kainate (50 μ M) to isolated horizontal cells invariably caused a large (0.3-5.0 nA) inward current accompanied by an increase in conductance and in membrane current noise (Figs 1b, c and 3a). This inward current was reduced at depolarized potentials and reversed to an outward current near 0 mV (0.57 ± 2.1 mV, mean $\pm$ s.e.m., $n = 7$), suggesting that kainate opens cation channels in the horizontal cell membrane¹⁶.

Kainate-gated currents were stable in the absence of dopamine. In the experiments reported here, the amount of kainate ejected was first adjusted to give a nonsaturating response (typically, 0.5-1.5 nA for cone horizontal cells, 1.0-2.5 nA for rod horizontal cells); repeated applications of the same quantity of kainate to a cell yielded responses that rarely varied in amplitude by more than 30%. In 10 cone horizontal cells that received only kainate, the mean response to the first application was 1.04 ± 0.15 nA ($n = 10$); subsequent responses averaged 0.90 ± 0.07 nA ($n = 18$). Following administration of a pulse of Ringer's containing dopamine (200 μ M), however, currents evoked by applications of kainate were markedly larger (Fig. 1b and Table 1). In 22 cone horizontal cells to which dopamine was applied, responses increased significantly ($P < 0.01$, two-tailed t -test) from 0.90 ± 0.21 nA ($n = 22$) to 1.95 ± 0.16 nA ($n = 82$). In 20 of these cells, dopamine increased kainate currents by more than 100% over control levels (mean maximal response 2.74 ± 0.34 nA) and changes of >10 -fold were occasionally observed.

That dopamine stimulates the production of cyclic AMP (cAMP) in fish retinas^{2,3}, and that dopamine has no significant

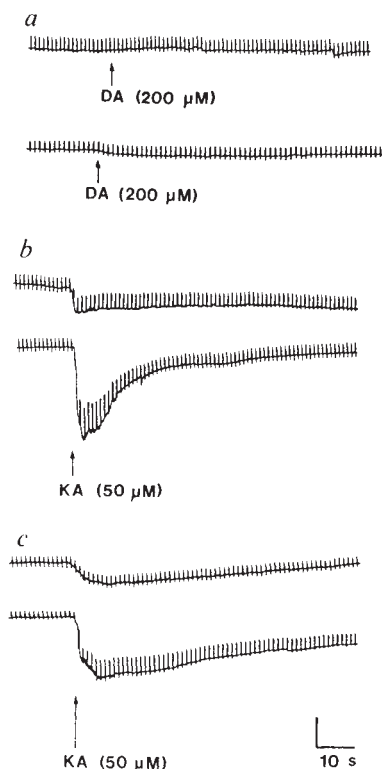


Fig. 1 Effects of dopamine (DA) on resting and kainate (KA)-gated conductances of cone horizontal cells. *a*, Voltage clamp records from two cells illustrating the lack of effect of dopamine on resting conductance. Membrane potential was clamped at -60 mV and 20 mV voltage pulses were applied throughout the record to monitor membrane conductance. Dopamine ($200 \mu\text{M}$ in Ringer's) was applied by pressure ejection in a pulse lasting 0.5 – 1.0 s at the times indicated by the arrows. Downward deflections denote inward transmembrane current. *b*, Voltage clamp records from another cell illustrating currents evoked by a pulse of Ringer's containing kainate ($50 \mu\text{M}$) before (upper trace) and 8 min after (lower trace) administration of $200 \mu\text{M}$ dopamine. Both kainate and dopamine were applied by pressure ejection as in *a*. *c*, Kainate currents in another cell before (upper) and 4 min after (lower trace) administration of 8 -bromo-cAMP ($500 \mu\text{M}$). Vertical calibration, 1.0 nA for *a* and *b*; 2.0 nA for *c*.

Methods. Horizontal cells were enzymatically and mechanically dissociated from white-perch retinas¹⁵, and maintained in culture for 1 – 10 days. During this time, four morphological classes of horizontal cells can be distinguished, three of which correspond to cone horizontal cells (unpublished data). Similar results were obtained from all three types of cone horizontal cell. The culture conditions used here do not affect the sensitivity of the dopamine-dependent adenylate cyclase in teleost horizontal cells (R. Van Buskirk and J.E.D., unpublished data), nor does this cyclase exhibit denervation supersensitivity²¹. Cells were voltage clamped using low resistance (5 – 14 M Ω) patch electrodes in the whole cell recording configuration^{17,18}, using a Dagan 8900 clamp amplifier with a 100 M Ω headstage (Dagan Corporation). In most experiments, the holding potential was set at -60 mV and membrane conductance assessed by monitoring the instantaneous currents evoked by 100 ms depolarizing 20 mV voltage steps. This potential range was chosen to minimize voltage-dependent sodium and potassium currents¹⁹, which were not otherwise blocked. The external medium was an oxygenated teleost Ringer's solution containing (mM): NaCl, 145 ; KCl, 4 ; CaCl_2 , 2.4 ; MgSO_4 , 1.7 ; MgCl_2 , 1.5 ; Na_2HPO_4 , 2.5 ; KH_2PO_4 , 0.4 ; glucose, 10 ; HEPES, 10 ; pH 7.6 . The pipette contained (mM): Potassium gluconate, 120 ; KCl, 4 ; MgCl_2 , 2 ; CaCl_2 , 1 ; EGTA, 11 ; HEPES, 10 ; pH 7.5 . Except for haloperidol, which was added directly to the bath, drugs were dissolved in Ringer's applied in brief (0.5 – 1.0 s) pulses by pressure ejection (100 – 400 nl) from small-bore triple-barrel pipettes placed 50 – $100 \mu\text{m}$ from the cell. All concentrations reported are those in the pipettes; we estimate that the concentration at the cell was lower by $\sim 25\%$ (ref. 13).

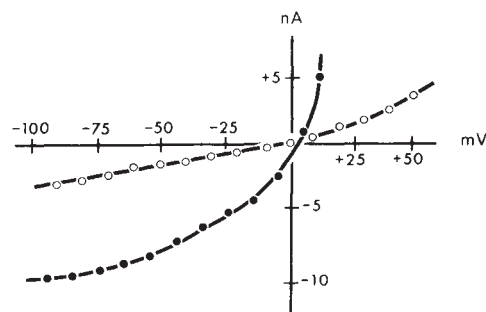


Fig. 2 Kainate current as a function of membrane potential for a cone horizontal cell that received only kainate ($\circ$) and for another cone horizontal cell following treatment with $500 \mu\text{M}$ 8 -bromo-cAMP ($\bullet$). Membrane potential was changed in a rampwise fashion (50 mV s^{-1}), and current-voltage relations for kainate were determined by subtracting currents recorded in the absence of kainate from currents recorded at the peak of a response to $50 \mu\text{M}$ kainate. Before exposure to 8 -bromo-cAMP, the response of the treated cell to kainate was 1.63 nA at a holding potential of -60 mV.

Table 1 Effects of test agents on responses of cone horizontal cells to $50 \mu\text{M}$ kainate

	Initial response (nA)	Response after drug (nA)	Time after administration of drug (s)	<i>n</i>
Control	1.04 ± 0.15	0.94 ± 0.08	183.5 ± 8.4	10
Dopamine ($200 \mu\text{M}$)	0.90 ± 0.21	$2.24 \pm 0.32^{**}$	195.2 ± 8.3	22
Dopamine ($200 \mu\text{M}$) + Haloperidol (20 – $50 \mu\text{M}$)	0.92 ± 0.16	0.90 ± 0.23	165.0 ± 6.0	7
8 -bromo-cAMP ($200 \mu\text{M}$) + Haloperidol (20 – $50 \mu\text{M}$)	0.82 ± 0.21	$2.37 \pm 0.67^*$	181.0 ± 11.2	10
Serotonin ($500 \mu\text{M}$)	1.40 ± 0.24	1.42 ± 0.36	154.3 ± 9.1	5

* $P < 0.05$, ** $P < 0.01$, two-tailed *t*-test.

Values given are mean $\pm$ s.e.m. For each cell tested, the amount of kainate ejected was first adjusted to give a nonsaturating response (0.3 – 2.5 nA). This dose was used in all subsequent applications of kainate to that cell. A pulse of Ringer's containing dopamine ($200 \mu\text{M}$), 8 -bromo-cAMP ($500 \mu\text{M}$), or serotonin ($500 \mu\text{M}$) was then applied, followed by repeated applications of the initial dose of kainate. For quantitative comparisons, the initial kainate response of each cell has been compared with a response obtained 2 – 4 min after application of the various test agents. For the control condition, where no drug was applied, a response obtained 2 – 4 min after the end of the initial response was chosen.

effects on membrane conductance, together suggest that cAMP mediates the enhancement phenomenon reported here. In support of this possibility, application to cone horizontal cells of a pulse of Ringer's containing $500 \mu\text{M}$ 8 -bromo-cAMP, a membrane permeable analogue of cAMP, also increases the amplitude of subsequent kainate responses (Fig. 1c and Table 1). The degree of enhancement produced by 8 -bromo-cAMP is similar to that produced by dopamine (mean initial response 0.82 ± 0.21 nA, $n = 10$; mean response after 8 -bromo-cAMP, 2.21 ± 0.34 nA, $n = 32$; $P < 0.05$). Neither 8 -bromo-cAMP nor dopamine cause any obvious changes in the reversal potential for kainate (Fig. 2), indicating that the increases in current reported here do not result from the activation of additional ionic conductances.

We have also examined whether dopamine affects the responses of cone horizontal cells to L-glutamate, which may be the transmitter released by photoreceptors^{10–14}. As for other species of fish^{16–20}, we found that perch horizontal cells were much less sensitive to glutamate than to kainate. Nevertheless,

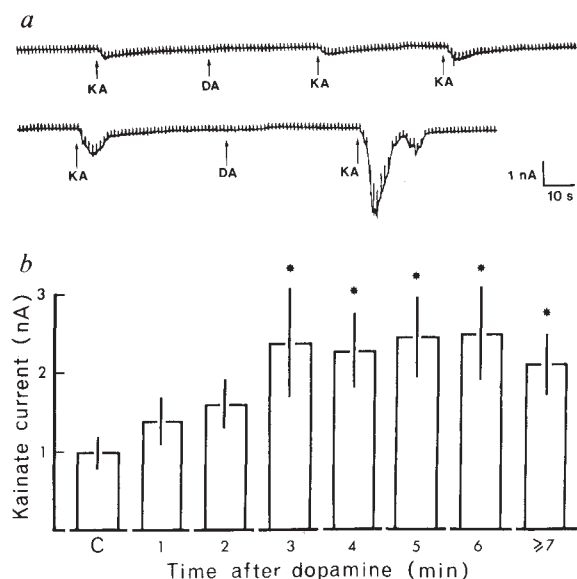


Fig. 3 Time course of the enhancement produced by dopamine. *a*, Voltage clamp record illustrating progressive increases in kainate (KA) currents after exposure to dopamine (DA). A second application of dopamine resulted in a further increase in response. Current fluctuations at the end of the largest response were not typical of responses seen after administration of dopamine. Method of drug application and other conventions as in Fig. 1. The record is continuous in time. *b*, Kainate-induced currents plotted as a function of time after application of dopamine. Data are from 12 cone horizontal cells which were all studied for ≥ 6 min. The experimental procedure was as described in Table 1, and all cells received only a single application of dopamine. Kainate responses were assigned to one-minute time bins depending on the interval between application of kainate and exposure to dopamine, with bin 1 containing responses obtained 0–1 min after application of dopamine, and so on. Bars, s.e.m.; asterisks, responses significantly greater than the mean control response (one-tailed *t*-test).

in eight cells studied, currents evoked by 100 μ M glutamate increased from 0.06 ± 0.03 nA to 1.15 ± 0.22 nA following exposure to 200 μ M dopamine ($P < 0.001$). The extent of enhancement was comparable to that seen with kainate.

The enhancing action of dopamine appeared to be specific to cone horizontal cells. The kainate-induced currents of rod horizontal cells, which lack dopaminergic innervation *in vivo*, failed to increase following treatment with dopamine (mean initial response 1.23 ± 0.54 nA, $n = 6$; mean response after treatment 0.98 ± 0.20 nA, $n = 8$) or with 8-bromo-cAMP (mean initial response 2.40 ± 0.79 nA, $n = 6$; mean response after treatment 2.22 ± 0.50 nA, $n = 10$). Furthermore, dopamine was ineffective when haloperidol (20–50 μ M) was added to the external medium, and application of serotonin (500 μ M) failed to alter the responsiveness of cone horizontal cells to kainate (Table 1). The evidence indicates that specific dopamine receptors on cone horizontal cells are involved in the enhancement of sensitivity to kainate.

Although the time course of the enhancement effect varied from cell to cell, it was generally consistent with the involvement of a second messenger system. As shown in Fig. 3, kainate-induced currents typically increased over the course of several minutes following the application of dopamine. In most cases, responses remained substantially elevated for as long as we could hold a cell (8–12 min). Additional applications of dopamine or 8-bromo-cAMP often resulted in additional enhancement, as was the case for the cell shown in Fig. 3*a*.

The results presented here demonstrate that dopamine, probably acting through a cAMP-dependent process, facilitates an ionic conductance gated by excitatory amino acids, possibly the same conductance used by the photoreceptor transmitter. Modu-

lation of voltage-dependent ion channels by second messenger systems is now well established; our findings suggest that channels gated by neurotransmitters may likewise be subject to modification by regulatory biochemical pathways. Our experiments do not reveal which properties of the amino acid-gated channels (affinity for ligand, number of functional channels, unitary conductance, kinetics) are altered by dopamine. Experiments at the single-channel level may resolve these issues.

An augmentation of amino-acid sensitivity may provide an explanation for the paradoxical observation that dopamine has no consistent effect on the membrane potential or resistance of isolated horizontal cells^{9,13}, but usually depolarizes horizontal cells in the intact retina^{4–6}. Under most experimental conditions, photoreceptors are thought to release a transmitter which tonically depolarizes horizontal cells. If dopamine facilitates the action of this transmitter, it would be expected to depolarize horizontal cells still further in the intact retina, when the transmitter is present, but to have no effect on isolated cells *in vitro*, where the transmitter is absent. Our results also suggest a possible explanation for the decrease in light responsiveness of horizontal cells observed following application of dopamine to the intact retina. Light reduces the flow of transmitter from the photoreceptors, thereby hyperpolarizing horizontal cells. If dopamine enhances the potency of the photoreceptor transmitter, as we have proposed, it may render light-induced reductions in transmitter release less effective and so decrease the amplitude of the responses of horizontal cells to light.

We thank David Corey and Barbara Barres for technical advice. Supported by NIH grants EY00824 and EY0588.

Received 28 July; accepted 27 November 1986.

1. Dowling, J. E. & Ehinger, B. *Proc. R. Soc. B* **201**, 7–26 (1978).
2. Van Buskirk, R. & Dowling, J. E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7825–7829 (1981).
3. O'Connor, P., Dorison, S., Watling, K. J. & Dowling, J. E. *J. Neurosci.* **6**, 1857–1865 (1986).
4. Hedden, W. L. & Dowling, J. E. *Proc. R. Soc. B* **201**, 27–55 (1978).
5. Mangel, S. C. & Dowling, J. E. *Science* **229**, 1107–1109 (1985).
6. Yang, X.-L., Tornqvist, K. & Dowling, J. E. *Invest. Ophthalm. vis. Sci.* **27**, 282 (1986).
7. Teranishi, T., Negishi, K. & Kato, S. *Nature* **301**, 243–246 (1983).
8. Piccolino, M., Neyton, J. & Gerschenfeld, H. M. *J. Neurosci.* **4**, 2477–2488 (1984).
9. Lasater, E. M., & Dowling, J. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3025–3029 (1985).
10. Murakami, M., Ohtsu, K. & Ohtsuka, T. *J. Physiol., Lond.* **227**, 889–913 (1972).
11. Slaughter, M. M. & Miller, R. F. *Science* **211**, 182–185 (1981).
12. Ishida, A. T. & Fain, G. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5890–5894 (1981).
13. Lasater, E. M. & Dowling, J. E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 936–940 (1982).
14. Miller, A. M. & Schwartz, E. A. *J. Physiol., Lond.* **334**, 325–349 (1983).
15. Dowling, J. E., Pak, M. & Lasater, E. M. *Brain Res.* **360**, 331–338 (1985).
16. Ishida, A. T. & Neyton, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1837–1841 (1985).
17. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
18. Marty, A. & Neher, E. in *Single Channel Recording* (eds Sakmann, B. & Neher, E.) 107–122 (Plenum, New York, 1983).
19. Lasater, E. M. *J. Neurophysiol.* **55**, 499–513 (1986).
20. Tachibana, M. *J. Physiol., Lond.* **358**, 153–167 (1985).
21. Watling, K. J., Parkinson, D. & Dowling, J. E. *Brain Res.* **253**, 334–336 (1982).

Spectral sensitivity of human cone photoreceptors

J. L. Schnapf*, T. W. Kraft & D. A. Baylor

Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305-5401, USA

The brain computes visual colour by analysing the relative excitations of three types of retinal cones¹. Each cone excitation is governed by a spectral sensitivity function which relates the amplitude of the neural response to wavelength at constant light intensity. The spectral sensitivities of human cones are not well characterized. We report measuring the sensitivities by recording electrical responses of human cones to stimuli of different wavelengths. Spectral sensitivities of 'green' and 'red' cones, determined over the entire visible region, show peaks near 530 and 560 nm respec-

* Present address: Department of Ophthalmology, University of California, San Francisco, California 94143, USA.

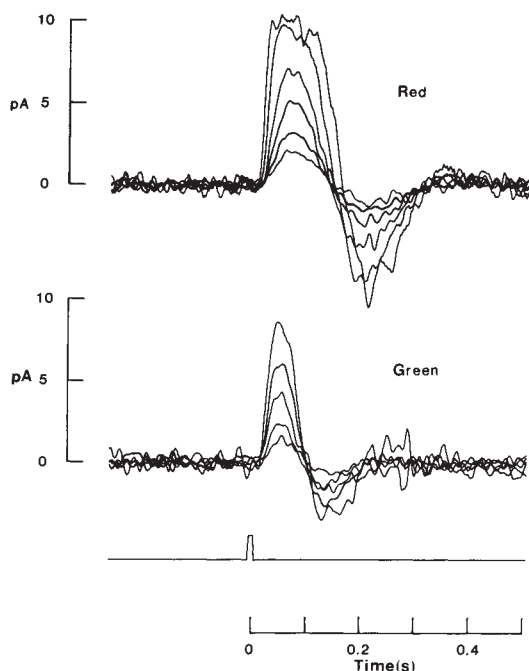


Fig. 1 Superimposed responses to flashes of increasing strength from a red (top) and a green cone (bottom). Outward change in membrane current is plotted as a function of the time after the flash. A flash monitor trace is also shown. Traces are averages of ≤ 11 responses. The light (500 nm) was unpolarized. Flash photon densities ranged from 252 to 16,220 photons μm^{-2} (top) and 252 to 2,542 photons μm^{-2} (bottom); bandwidth 0–50 Hz.

tively, and are remarkably similar to those of the old-world monkey *Macaca fascicularis*^{2,3}. They satisfactorily predict the photopic luminosity function, a measure of the sensitivity of cone-mediated human vision to light of different wavelengths. The kinetics of the light responses of human cones also appeared similar to those of macaque cones: the time to peak response to a dim flash was 50–100 ms and there was a characteristic undershoot during recovery.

The retina was obtained from the eye of a 59-year-old man with a tumour in the anterior chamber. Acuity of the eye, tested the day before enucleation, was 20/30. The colour vision in the other eye was normal when tested by Ishihara plates after surgery. Following enucleation, the posterior pole of the eye was removed under dim red light and subsequent manipulations were performed under infrared light using infrared/visible image converters. The retina was isolated, chopped into fine pieces

~100 μm across, and transferred to a recording chamber containing a bicarbonate-buffered Locke's solution⁴ supplemented by amino acids and vitamins from Basal Medium Eagle (Gibco). The chamber was continuously perfused and kept at a temperature of 37 °C. Photocurrent was collected from single cone outer segments with a glass micropipette⁵. Spectral sensitivity was measured at wavelengths between 381 and 780 nm by applying flashes of monochromatic light (10 nm band width) and finding the reciprocal flash strength (in photons⁻¹ μm^2) needed to evoke a response of a fixed amplitude^{2,4,6}. The quantum sensitivity at each wavelength was determined relative to a reference sensitivity at 500 nm, which was measured frequently to avoid errors due to slow changes in the state of the cell. A complete experiment required about one hour of recording and bleached ~3% of the pigment in the outer segment³. Since the light path for transverse illumination of the outer segment was relatively short (<4 μm) the measured spectral sensitivity should be proportional to the molecular extinction coefficient of the visual pigment molecules in the cone, assuming that an absorbed photon excites with an efficiency that does not depend upon wavelength.

Photocurrent was recorded from two green cones and five red cones, identified by the quantal sensitivity at 660 nm relative to that at 500 nm: 0.007 (green) and 0.1 (red). Recordings were also obtained from a few rods but their responses were small and insensitive, perhaps because of exposure to the surgical lights. Cone responses to light flashes of increasing strength are illustrated in Fig. 1. The stimuli evoked graded outward-going photocurrents that returned to the baseline level with an undershoot. The time to the peak of the response to a dim flash was 50–100 ms in different cells. These kinetic properties are similar to those of macaque cones^{2,3}. As the flash strength was increased, the photocurrent amplitude grew according to the exponential saturation function

$$r/r_m = (1 - \exp(-ki))$$

where r is the peak amplitude of the response, r_m the maximal amplitude, i the flash photon density and k a constant^{3,7}. The value of r_m was ~10 pA, and the average value of k for unpolarized light at the wavelength of maximal sensitivity was $\sim 10^{-3}$ photons⁻¹ μm^2 , corresponding to a half-saturating photon density of ~700 photons μm^{-2} .

The points in Fig. 2 show the spectral sensitivity of one green cone and the average sensitivity of the five red cones. (The spectral sensitivity of the second green cone was not completely determined.) The peak sensitivities were at ~530 nm (green) and ~560 nm (red). The smooth curves, which fit the average spectra of macaque cones³, are very similar to the human spectra. Absorption spectra obtained from single human cones by micro-

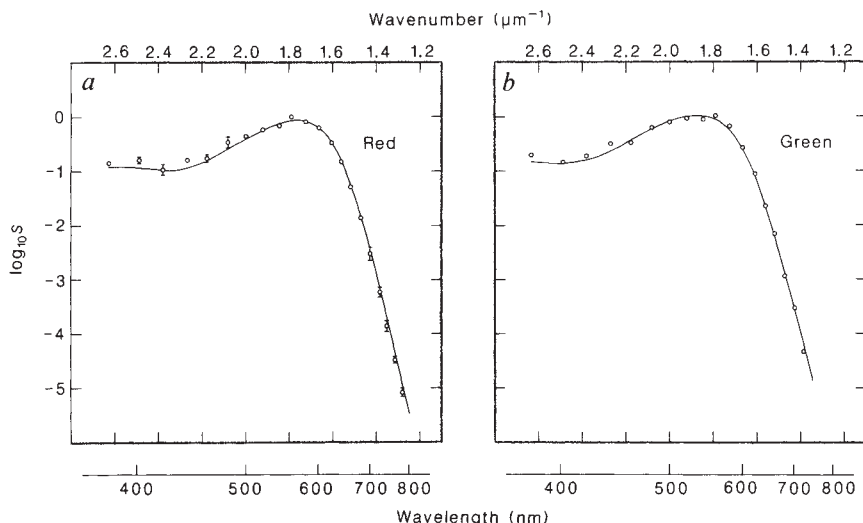


Fig. 2 Spectral sensitivity of human cones. Symbols show the normalized sensitivity of the average of five red cones (a) and a single green cone (b) as a function of wavenumber, on semi-log axes. Bars, standard deviation in a; the 780 nm point is an average over only 4 cones. Wavelength scale below. The smooth curves fit the average spectra of the red and green cones of the monkey *Macaca fascicularis*³.

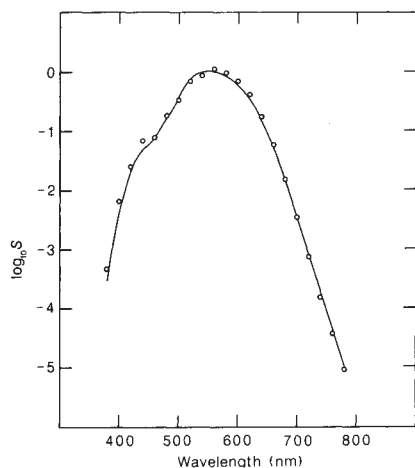


Fig. 3 Prediction of the photopic luminosity function from spectral sensitivities of the red and green cones. Smooth curve, photopic luminosity function given by Vos¹⁵, obtained from tabulations in Wyszecki and Stiles¹³ and recalculated on a quantum basis; points, weighted sum of the red and green cone spectral sensitivities, corrected for the effects of pre-retinal absorption and pigment self-screening (see text). The weighting of the red and green cone contributions was 1:0.59.

spectrophotometry⁸ show roughly the same shape and position but are somewhat broader.

The flatness of the peaks of the spectra made it difficult to determine accurately the wavelength of maximum sensitivity. The linear decline of the logarithm of sensitivity ($\log S$) at low wavenumber², however, proved useful for characterizing the position of spectra on the wavenumber axis. Thus the position can be specified by the wavenumber at which a straight line fitted to the average spectrum at low wavenumber intersects $\log_{10} S = 0$ (spectra normalized to the value at 560 nm). When the line was fitted to the average red cone spectrum over the points 680–780 nm, the wavelength of intersection was 633 nm. This compares to a value of 634 nm for the red cones of the macaque, determined in the same way. The intersection of a line fitted to the human green cone spectrum at wavelengths 640–720 nm was 600 nm, compared to 598 nm for the macaque green cones. Within experimental error the human and macaque cone spectra appear identical.

Microspectrophotometric spectra from single human cones of a given type vary considerably, suggesting the existence of subtypes⁸. To investigate variability among the 5 red cones, the straight line fitted to the average red spectrum was shifted along the abscissa to reach the best fit for each measured spectrum by the least squares criterion. The position of a spectrum was again characterized by the wavenumber at which the line crossed the $\log_{10} S = 0$ level. Using this method, the standard deviation of the positions, expressed as a wavelength interval, was found to be 0.9 nm. The observed variability probably resulted from experimental error so that in this sample of cones there was no evidence for subtypes. A similar dispersion was noted in a sample of 15 red cones from 6 male macaques, which had a standard deviation of 1.0 nm (ref. 3).

The descent of sensitivity at low wavenumber was steeper for the red cones than for the green. The asymptotic value of the slope was calculated by linear regression on individual points from all cells at values of $\log_{10} S$ that were less than -2.9 . The slope in $\log_{10}$ units μm was (slope $\pm$ s.e.m.) 17.5 ± 0.5 for the red cones and 16.2 ± 0.8 for the green. These compare with values of 17.2 ± 0.2 and 15.9 ± 0.2 obtained from the cones of the macaque³. The ratio of excitation in the red and green cones increases as the wavelength approaches 700 nm but then declines because of the difference in the long-wavelength slopes. This

explains the 'paradoxical hue shift'⁹ in which light at very long wavelengths resembles that at characteristic shorter wavelengths². The asymptotic red slope was close to the value of 17.4 $\log_{10}$ units μm obtained for the long wavelength descent of sensitivity of human cone vision^{10,11} which is probably mediated by only the red cones.

Only the red and green cones are thought to contribute to luminosity¹², a psychophysical quantity roughly related to brightness¹³. We have examined this notion by attempting to predict, as a simple sum of the red and green curves, the Commission Internationale de l'Eclairage (CIE) photopic luminosity function modified by Judd¹⁴ and further modified by Vos¹⁵. This function describes the luminosity of light of different wavelengths when only the cones are active. The values tabulated by Wyszecki and Stiles¹³ were recalculated on a quantum basis and plotted as the smooth curve in Fig. 3.

For comparison with psychophysical results the measured cone spectral sensitivity functions were modified to take into account self-screening of the visual pigment and preretinal screening by the lens and macular pigment. To correct for self-screening, we applied Beer's law with a maximum axial pigment density of 0.27 (ref. 3). Preretinal absorption was calculated from published tabulations (ref. 13, Table 1 (2.4.6), column b, and Table 2 (2.4.6)). Spectra modified in this way describe the sensitivity of the red and green cones to light incident on the cornea.

An iterative computer program determined the relative weighting of the modified cone spectra that gave the best fit to the experimentally-observed luminosity function. Fitting was performed only at wavelengths ≥ 500 nm to reduce errors associated with uncertainties about the magnitudes of preretinal screening. The best fit was obtained when the red and green cone sensitivities were scaled in the ratio 1:0.59. The weighted sum of the sensitivities is plotted as the symbols in Fig. 3 and provides a reasonable approximation to the luminosity function. A similar weighting of the red and green cone contributions (1:0.64) was derived from psychophysical experiments¹². The results in Fig. 3 are consistent with the idea that only the red and green cones contribute to luminosity in cone vision.

Previous work on macaque cones yielded spectral sensitivities in good agreement with psychophysical estimates of human cone sensitivities³. The present work supports the supposition that macaque and human cones have very similar or identical spectral properties. Recently the amino-acid sequences of human cone pigments have been determined¹⁶. Given the similarity of the human and macaque spectra, a comparison of the sequences may help to identify the regions responsible for spectral tuning. Spectral sensitivities determined by the method described here should prove useful in formulating and testing quantitative models of colour vision.

We thank Drs P. Egbert and M. Marmor for invaluable cooperation, Professors R. Boynton and L. Stryer for discussion and Mr R. Schneeveis for technical assistance. Supported by grant EY05750 from the National Eye Institute, and NIH Research Service Award NS07158.

Received 22 September; accepted 1 December 1986.

- Boynton, R. M. *Human Color Vision* (Holt, Rinehart & Winston, New York, 1979).
- Nunn, B. J., Schnapf, J. L. & Baylor, D. A. *Nature* **309**, 264–266 (1984).
- Baylor, D. A., Nunn, B. J. & Schnapf, J. L. *J. Physiol. Lond.* (in the press).
- Baylor, D. A., Nunn, B. J. & Schnapf, J. L. *J. Physiol. Lond.* **357**, 575–607 (1984).
- Baylor, D. A., Lamb, T. D. & Yau, K.-W. *J. Physiol. Lond.* **288**, 589–611 (1979).
- Naka, K. I. & Rushton, W. A. H. *J. Physiol. Lond.* **185**, 536–555 (1966).
- Lamb, T. D., McNaughton, P. A. & Yau, K.-W. *J. Physiol. Lond.* **319**, 463–496 (1981).
- Dartnall, H. J. A., Bowmaker, J. K. & Mollon, J. D. *Proc. R. Soc. B220*, 115–130 (1983).
- Brindley, G. S. *J. Physiol. Lond.* **130**, 35–44 (1955).
- Goodeve, C. F. *Proc. R. Soc. A155*, 664–683 (1936).
- Griffin, D. R., Hubbard, R. & Wald, G. *J. opt. Soc. Am.* **37**, 546–554 (1947).
- Smith, V. C. & Pokorny, J. *Vision Res.* **15**, 161–171 (1975).
- Wyszecki, G. & Stiles, W. S. (eds) *Color Science* (Wiley, New York, 1982).
- Judd, D. B. *CIE Proc.* **1**, 11 (1951).
- Vos, J. J. *Col. Res. Appl.* **3**, 125 (1978).
- Nathans, J., Thomas, D. & Hogness, D. S. *Science* **232**, 193–202 (1986).

A cyclic nucleotide-gated conductance in olfactory receptor cilia

Tadashi Nakamura & Geoffrey H. Gold

Department of Physiology, Yale University School of Medicine,
333 Cedar Street, New Haven, Connecticut 06510, USA

Olfactory transduction is thought to be initiated by the binding of odorants to specific receptor proteins in the cilia of olfactory receptor cells (reviewed in refs 1–3). The mechanism by which odorant binding could initiate membrane depolarization is unknown, but the recent discovery of an odorant-stimulated adenylyl cyclase in purified olfactory cilia^{4,5} suggests that cyclic AMP may serve as an intracellular messenger for olfactory transduction. If so, then there might be a conductance in the ciliary plasma membrane which is controlled by cAMP. Here we report that excised patches of ciliary plasma membrane, obtained from dissociated receptor cells, contain a conductance which is gated directly by cAMP. This conductance resembles the cyclic GMP-gated conductance that mediates phototransduction in rod and cone outer segments^{6,7}, but differs in that it is activated by both cAMP and cGMP. Our data provide a mechanistic basis by which an odorant-stimulated increase in cyclic nucleotide concentration could lead to an increase in membrane conductance and therefore, to membrane depolarization. These data suggest a remarkable similarity between the mechanisms of olfactory and visual transduction and indicate considerable conservation of sensory transduction mechanisms.

We examined the conductance of the ciliary plasma membrane using the patch-clamp technique⁸. First, a gigohm seal was obtained on a cilium of a dissociated olfactory receptor cell (Fig. 1). Because olfactory cilia are only $\sim 0.25 \mu\text{m}$ in diameter⁹, small patch pipette tips were necessary to obtain gigohm seals (bubble number (ref. 10) < 1.6 , pipette resistance 30–50 M Ω). The membrane patch was excised with a gentle tap on the micromanipulator, exposing the cytoplasmic surface of the membrane to the bath solution. In a few cases, the cyclic nucleotide-gated conductance was observed immediately after patch excision, but more often the excised patch sealed over to form a closed vesicle. Once a vesicle formed, the cyclic nucleotide-gated conductance could not be observed until the outer half of the vesicle was ruptured by touching the tip of the patch pipette to a Sylgard surface¹¹. We characterize membrane conductance by measuring the current-voltage relationships of the excised patches. An investigation of the single channel events which underlie these currents will be presented elsewhere.

Figure 2a shows current-voltage relationships obtained from an excised patch of ciliary plasma membrane in the presence and absence of cAMP. In a total of 46 patches we found that cAMP induced a reversible increase in membrane conductance which occurs in the absence of added nucleotide triphosphates,

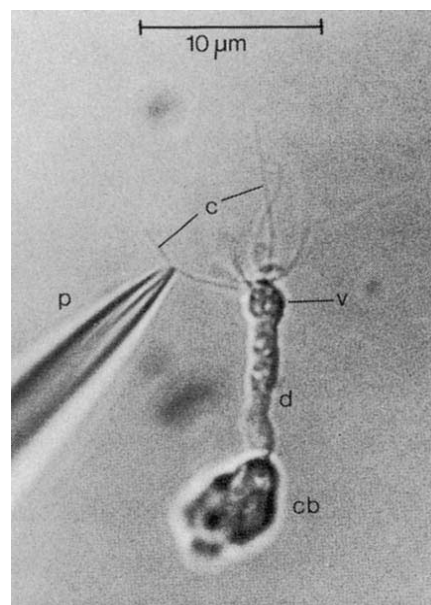


Fig. 1 Light micrograph of a dissociated olfactory receptor showing the location of the patch pipette (p) during gigohm seal formation on a cilium (c). Olfactory receptors were identified by their characteristic morphology³¹; extending from the cell body (cb), is a single dendrite (d), which terminates in a small swelling, termed the olfactory vesicle (v). Several cilia are just visible; the low contrast of the cilia is due to their small size ($\sim 2,500 \text{ \AA}$ diameter). **Methods.** Solitary olfactory receptor cells were obtained from the olfactory epithelia of toads (*Bufo marinus*) by enzymatic dissociation with papain¹⁸. Our preparations contained small pieces of olfactory epithelium as well as solitary receptors, supporting cells and respiratory cells. Receptors could readily be identified by their characteristic morphology³¹. Gigohm seals were obtained with fire-polished borosilicate glass microelectrodes. The patch current was measured with a conventional current-to-voltage converter circuit (Yale Mark V). The current signal was low-pass filtered before being digitized at 1 kHz. A laboratory computer system (INDEC) was used to generate command voltage ramps to obtain the $I-V$ relationships (ramp rate = 0.5 mV ms^{-1}). Voltage ramps alternated between ascending and descending directions and were averaged to obtain the single traces shown in Figs 2 and 3. Solutions: Ringer's solution (mM): NaCl, 110; KCl, 2.5; MgCl_2 , 1.6; CaCl_2 , 1; HEPES, 5; pH 7.4. Ringer's solution with $p\text{Ca} \sim 7$: NaCl, 95; KCl, 2.5; MgCl_2 , 1.6; CaCl_2 , 0.25; EGTA, 0.5; pH 7.4. Pseudointracellular solution: NaCl, 12.5; KCl, 85; MgCl_2 , 1.6; CaCl_2 , 0.25; EGTA, 0.5; HEPES, 5; pH 7.4. Divalent cation-free solution^{14–15}: NaCl, 118; EDTA, 0.1; EGTA, 0.1; HEPES, 5; pH 7.4.

and must therefore result from direct gating of the conductance by cAMP rather than from cAMP-stimulated protein phosphorylation. Direct gating of a conductance by cyclic nucleotides has been observed previously, although only in excised patches from the outer segments of rod and cone photoreceptors^{6,7}.

The cAMP-activated conductance is nearly ohmic at negative potentials, but is rectifying at positive potentials. This rectification is seen only in the presence of normal levels of Ca^{2+} and Mg^{2+} ; removal of divalent cations from both sides of the membrane nearly abolishes the rectification and greatly increases the membrane conductance at 0 mV (compare Figs. 2 and 3). Thus the ciliary conductance appears to exhibit voltage-dependent blockage by divalent cations, like the cGMP-gated conductance of the rod outer segment^{12–15}.

With Ringer's solution in the patch pipette and a pseudointracellular solution in the bath (Fig. 2b), the reversal potential for the cAMP-gated conductance was close to zero ($-5 \pm 3 \text{ mV}$; mean $\pm$ s.d., number of observations (N) = 7). This is quite similar to the reversal potential for the odorant-stimulated con-

Table 1 Relative affinities of the cyclic nucleotide-gated conductance for different cyclic nucleotides

cAMP	cGMP	cCMP
3.0 (1.3)	1.6 (1.1)	50 (1.4)
2.4 (1.8)	1.6 (1.7)	64 (1.8)
3.3 (1.6)	2.1 (1.6)	—
—	0.7 (1.4)	—
4.9 (1.4)	—	—
37 (1.5)	20 (1.6)	—
—	20 (1.9)	—

Each line records data obtained from a single patch. Affinity is expressed as $K_{1/2}$ in units of μM . Figures in parentheses are the value of the Hill coefficient used to fit the Hill equation to the experimental data.

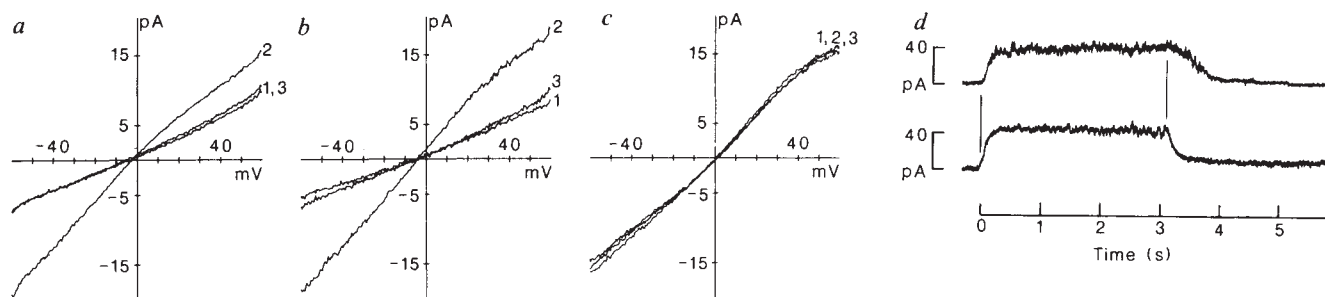


Fig. 2 *a-c*, Current-voltage relationships observed in the presence and absence of cyclic nucleotides. *a*, Traces were measured (1) before, (2) during and (3) after the bath application of 1 mM cAMP. The pipette contained Ringer's solution; the bath contained Ringer's solution with $pCa \sim 7$. Because there was no gradient for monovalent cations, the reversal potential was ~ 0 . *b*, Traces 1, 2 and 3 were measured before, during and after the bath application of 1 mM cAMP in a pseudo-intracellular solution. The pipette contained Ringer's solution. The reversal potential for this patch was 0 mV; the reversal potential of the cyclic nucleotide-gated conductance was always referred to the reversal potential of the leak conductance, because the latter varied due to uncompensated voltage offsets (always < 5 mV). *c*, Traces 1-3 were obtained in the presence of 1 mM cAMP, 1 mM cGMP and 1 mM of both cAMP and cGMP. The leak trace (current-voltage relationship in the absence of added cyclic nucleotides) was subtracted to illustrate better the conductance change induced by the cyclic nucleotides. The small upward inflection in the current-voltage curves around 0 mV was not seen in all patches. The pipette contained Ringer's solution; the bath contained Ringer's solution with $pCa \sim 7$. The data in *a-c* were each obtained from different patches and indicate the variability among patches in the presence of divalent cations. Excised patches were more stable in the absence of divalent cations, possibly because of the reduced ability of membrane patches to re-form a closed vesicle. Consequently, all data, except those shown in Fig. 2*a-c*, were obtained with a divalent cation-free solution^{14,15} on both sides of the membrane. *d*, Upper trace, time course of the conductance change following step changes in cAMP concentration from 0 to 10 μM and back to 0 μM ; lower trace, the time course of the patch current following step changes in bath Na concentration from 0 to 110 mM and back to 0 mM, all containing 10 μM cAMP (choline was substituted for Na). The solution changes were performed by manually rotating the perfusion selector valve at a predetermined time after data acquisition was initiated. Due to the variability in the manual control of solution changes, the timing of the solution changes is only accurate to within a few hundred milliseconds. The membrane potential was fixed at 30 mV and the divalent cation-free solution was present on both sides of the membrane.

ductance¹⁶⁻¹⁸, consistent with the hypothesis that this cAMP-gated conductance mediates olfactory transduction. The small absolute value of the reversal potential indicates that this conductance does not select between sodium and potassium ions, also a characteristic of the cGMP-gated conductances of rod and cone outer segments^{6,7}.

Figure 2*d* (upper trace) shows the time course of the conductance change following step changes in cAMP concentration from 0 to 10 μM and back to 0 μM . The time course of the solution change is inferred from the response to identically timed step changes in sodium ion concentration in the presence of 10 μM cAMP (lower trace). The similarity between the two traces indicates that the ciliary conductance responds to cyclic nucleotide concentration changes in less than a few hundred milliseconds, and therefore is fast enough to mediate olfactory transduction¹⁹.

Because the outer segment conductances of rods and cones are highly specific for cGMP, we investigated the sensitivity of the ciliary conductance to other cyclic nucleotides. Figure 2*c* shows the conductance increase produced by the bath application of 1 mM cAMP, 1 mM cGMP, and 1 mM of both cAMP and cGMP. Both cAMP and cGMP fully activate the ciliary conductance. As the same maximal conductance is induced by 1 mM of cAMP or cGMP alone or both together, there must be a single conductance which is gated by both cyclic nucleotides, rather than separate cAMP- and cGMP-gated conductances. In other patches, we found that 1 mM cyclic cytidine monophosphate (cCMP; the only other cyclic nucleotide tested) induced the same maximal conductance increase as that produced by cAMP or cGMP. The gating of the ciliary conductance is, nevertheless, highly specific for nucleotide 3',5'-cyclic monophosphates because 1 mM 2',3'-cAMP, 2',3'-cGMP, 5'-AMP and 5'-GMP had no significant effect. The responsiveness of the patches to cAMP and cCMP could not have been caused by contamination of our reagents with cGMP, because we were able to demonstrate the high selectivity of the rod outer segment conductance for cGMP over both cAMP and cCMP (data not shown).

The ligand specificity of the ciliary conductance was determined by measuring its concentration dependence for several cyclic nucleotides (Fig. 3), obtained from the slope, at 0 mV, of the $I-V$ relationships, which were recorded in various ligand concentrations (Fig. 3*a*). The normalized patch conductance was plotted as a function of cyclic nucleotide concentration and fitted to the Hill equation (Fig. 3*b*) to obtain values of $K_{1/2}$ and n , the cooperativity factor (Table 1). In all four patches on which both cAMP and cGMP were tested, the affinity of the conductance was 1.7 ± 0.2 (mean $\pm$ s.d.) times higher for cGMP than for cAMP. Thus, the gating mechanism has a small but significant preference for cGMP over cAMP. In the first 5 patches in Table 1, the $K_{1/2}$ s for cAMP and cGMP are in the micromolar range, whereas in the last two patches the $K_{1/2}$ s were ~ 10 times higher. This apparently bimodal distribution is suggestive of an intrinsic heterogeneity in olfactory receptors which might reflect different classes of receptors^{17,20}, or different stages in the development of olfactory receptors²¹, which continually turn over in the adult animal²². In the two patches in which cCMP was also tested, the affinity for cCMP was ~ 40 times lower than for cGMP. Thus, although the conductance can be fully activated by 1 mM cCMP, cCMP seems less likely to control this conductance *in vivo*. The cooperativity factor used to fit the Hill equation to the data ($n = 1.5 \pm 0.2$; mean $\pm$ s.d., $N = 13$) indicates positive cooperativity of cyclic nucleotide molecules in activating the conductance; positive cooperativity has also been observed for the rod and cone conductances, although there the reported values of n are somewhat larger^{6,7,14,15}. Although the current-voltage relationship in Fig. 3*a* is virtually linear at saturating cyclic nucleotide concentrations, at sub-saturating concentrations the current-voltage relationship is noticeably curvilinear. The upward curvature indicates that membrane depolarization increases opening of the channel by cyclic nucleotides, suggesting either that the cyclic nucleotide binding sites are located within the bilayer lipid membrane, or that membrane potential can influence the affinity of the binding sites for cyclic nucleotides. This phenomenon has also been observed in the rod outer segment conductance^{14,15}.

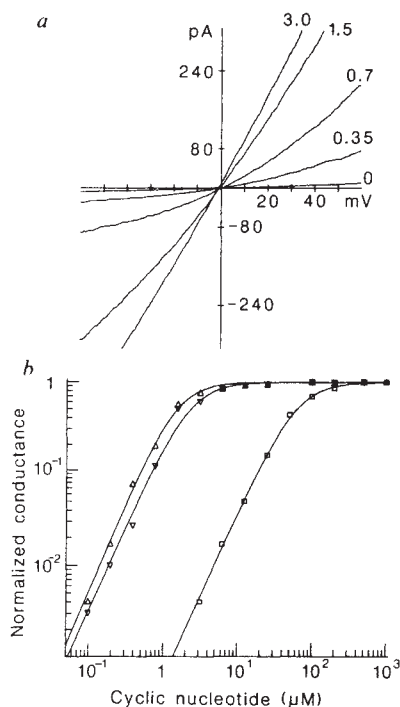


Fig. 3 Concentration dependence of the cyclic nucleotide-gated conductance measured with the divalent cation-free solution on both sides of the membrane. *a*, *I*-*V* curves measured in the presence of several concentrations of cGMP (concentration (μM) indicated to the right of each trace). *b*, Normalized conductance plotted as a function of cyclic nucleotide concentration: Δ, cGMP; ▽, cAMP; □, cCMP (all data obtained from a single patch). The *I*-*V* curves were analysed as follows: first the leak conductance was subtracted from all curves and then the saturating *I*-*V* curve was linearly scaled until the slope at 0 mV superimposed on each of the subsaturating *I*-*V* curves. The scaling factors are plotted as a function of concentration. The Hill equation, $g = C^n / (C^n + K_{1/2}^n)$, where C = cyclic nucleotide concentration, and n = the Hill coefficient, was fitted by eye to the data points and is plotted as a continuous line for each cyclic nucleotide. The values of $K_{1/2}$ and n used for these curves were 1.6 μM and 1.7 for cGMP, 2.4 μM and 1.8 for cAMP and 64 μM and 1.8 for cCMP.

The presence of the cyclic nucleotide-gated conductance in olfactory receptor cilia is consistent with evidence that the cilia are the site of olfactory transduction^{23,24}. However, we have also observed a cAMP- and cGMP-gated conductance in patches excised from the dendrite and cell body regions. The *I*-*V* relationships observed in the dendritic and somatic patches were similar to those observed in the ciliary patches, indicating that the same, or a similar, conductance was present in the cilia, dendrite and soma. We do not know whether the presence of this conductance throughout the cell occurs *in vivo* or is a consequence of tissue dissociation. Its presence in the distal dendrite might be important for generating current in response to cyclic nucleotides that may be produced in the olfactory vesicle, or that may leak out of the bases of the cilia. With high odorant concentrations, the dendritic conductance might extend the dynamic range of olfactory receptors by increasing the total membrane current under conditions in which the ciliary currents are saturated.

The demonstration of a cyclic nucleotide-gated conductance in olfactory receptor cilia, together with recent reports of an odorant-stimulated adenylate cyclase in olfactory cilia^{4,5} are strong evidence that olfactory transduction is mediated by cyclic nucleotides. Specifically, we propose that odorant-stimulated cyclase activity produces an increase in cyclic nucleotide concentration which, by a rapid and direct effect on the conductance

reported here, initiates the depolarizing response to odorants. Further elaboration of this model may be needed to incorporate odorant-stimulated polyphosphoinositide turnover²⁵, or to explain the response to odorants which do not appear to stimulate the adenylate cyclase²⁶. Furthermore, this cyclic nucleotide model of olfactory transduction does not exclude the possibility that certain odorants may gate channels in the ciliary plasma membrane directly^{27,28}.

The ciliary axoneme in vertebrate photoreceptors has been interpreted as evidence for an evolutionary relationship between visual and olfactory receptor cells²⁹. The similarities between the cyclic nucleotide-gated conductances of the olfactory receptor and of the rod and cone photoreceptors support this notion, suggesting considerable conservation of sensory transduction mechanisms. Therefore, the differences between the photoreceptor and olfactory receptor conductances may reflect differences between optimal strategies which have evolved for the detection of olfactory and visual stimuli. For example, the ciliary conductance has an affinity for cGMP ~20 times higher than that of the photoreceptor conductance, perhaps because light causes a decrease in cGMP concentration, whereas odorants increase cyclic nucleotide concentration. Consequently, olfactory receptors could achieve higher sensitivity, without additional metabolic cost, simply by using a higher affinity conductance. It is not known why the conductance lacks specificity for a single cyclic nucleotide. Possibly, it is optimized to detect all products of the odorant-stimulated cyclase (or cyclases), which may produce both cGMP and cAMP *in vivo*. A lack of substrate selectivity in a single species of enzyme could increase reaction velocity³⁰, and contribute to the receptor's high sensitivity to odorants. Alternatively, distinct adenylate and guanylate cyclases could provide a single receptor cell with separate biochemical pathways for transduction. Future biochemical studies may distinguish between these alternatives.

We thank G. M. Shepherd for stimulating our interest in this problem. We also thank W. K. Chandler, B. Forbush, L. M. Masukawa and E. N. Pugh for critical reviews of the manuscript. This work was supported by NIH grant EY03955.

Received 22 September; accepted 27 November 1986.

1. Rhein, L. D. & Cagan, R. H. in *Biochemistry of Taste and Olfaction* (eds Cagan, R. H. & Kare, M. R.) 47-65 (Academic, New York, 1981).
2. Getchell, T. V. *Physiol. Rev.* **66**, 772-818 (1986).
3. Lancet, D. A. *Rev. Neurosci.* **9**, 329-355 (1986).
4. Pace, U., Hanski, E., Salomon, Y. & Lancet, D. *Nature* **316**, 255-258 (1985).
5. Sklar, P. B., Anholt, R. H. & Snyder, S. H. *J. biol. Chem.* **261**, 15538-15543 (1986).
6. Fesenko, E. E., Kolesnikov, S. S. & Lyubarsky, A. L. *Nature* **313**, 310-313 (1985).
7. Haynes, L. W. & Yau, K.-W. *Nature* **317**, 61-64 (1985).
8. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
9. Reese, T. S. *J. Cell Biol.* **25**, 209-230 (1965).
10. Corey, D. P. & Stevens, C. F. in *Single Channel Recording* (eds Sakmann, B. & Neher, E.) 53-68 (Plenum, New York, 1983).
11. Sakmann, B. & Neher, E. in *Single Channel Recording* (eds Sakmann, B. & Neher, E.) 37-51 (Plenum, New York, 1983).
12. Matthews, G. *Soc. Neurosci. Abstr.* **11**, 1130 (1985).
13. Yau, K.-W. & Haynes, L. W. *Biophys. J.* **49**, 33a (1986).
14. Haynes, L. W., Kay, A. R. & Yau, K.-W. *Nature* **321**, 66-70 (1986).
15. Zimmerman, A. L. & Baylor, D. A. *Nature* **321**, 70-72 (1986).
16. Trotter, D. & MacLeod, P. *Brain Res.* **268**, 225-237 (1983).
17. Suzuki, N. in *Food Intake and Chemical Senses* (eds Katsuki, Y., Sato, M., Takagi, S. F. & Oomura, Y.) 13-22 (University of Tokyo Press, 1977).
18. Anderson, P. A. V. & Hamilton, K. A. *Neuroscience* (in the press).
19. Getchell, T. V., Heck, G. L. & DeSimone, J. A. *Biophys. J.* **29**, 397-412 (1980).
20. Getchell, T. V. & Shepherd, G. M. *J. Physiol., Lond.* **282**, 541-560 (1978).
21. Masukawa, L. M., Hedlund, B. & Shepherd, G. M. *J. Neurosci.* **5**, 136-141 (1985).
22. Takagi, S. F. in *Handbook of Sensory Physiology* Vol. 4 (ed. Beidler, L. M.) 75-94 (Springer, New York, 1971).
23. Rhein, L. D. & Cagan, R. H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4412-4416 (1980).
24. Adamek, G. D., Gesteland, R. C., Mair, R. G. & Oakley, B. *Brain Res.* **310**, 87-97 (1984).
25. Huque, T. & Bruch, R. C. *Biochem. biophys. Res. Commun.* **137**, 36-42 (1986).
26. Sklar, P. B., Anholt, R. H. & Snyder, S. H. *J. Neurosci. Abstr.* **12**, 1178 (1986).
27. Vodyanov, V. & Murphy, R. B. *Science* **220**, 717-719 (1983).
28. Fesenko, E. E., Novoselov, V. I., Pervukin, G. Y. & Fesenko, N. K. *Biochim. biophys. Acta* **466**, 347-356 (1977).
29. Vinnikov, J. A. *Cold Spring Harb. Symp. quant. Biol.* **30**, 293-300 (1965).
30. Jencks, W. P. in *Chemical Recognition in Biology* (eds Chapeville, F. & Haenni, A.-L.) 3-25 (Springer, New York, 1980).
31. Graziadei, P. P. C. in *Handbook of Sensory Physiology* Vol. 4 (ed. Beidler, L. M.) 27-58 (Springer, New York, 1971).

The GTP-binding protein, G_o , regulates neuronal calcium channels

J. Hescheler*, W. Rosenthal†, W. Trautwein* & G. Schultz†

* II Physiologisches Institut, Universität des Saarlandes, D-6650 Homburg/Saar, FRG

† Institut für Pharmakologie, Freie Universität Berlin, D-1000 Berlin 33, FRG

In neuronal cells, opioid peptides and opiates inhibit neurotransmitter release, which is a calcium-dependent process¹. They also inhibit adenylyl cyclase², presumably via the membrane signal-transducing component, G_i , a guanine nucleotide-binding protein (G-protein)^{3,4}. No causal relationship between these two events has yet been demonstrated. Besides G_i , membranes of neuronal tissues contain large amounts of G_o , a G-protein with unknown function^{5,6}. Both G-proteins are heterotrimers consisting of α -, β - and γ -subunits⁷; the α -subunits can be ADP-ribosylated by an exotoxin from *Bordetella pertussis* (PT), which modification inhibits receptor-mediated activation of the G-protein⁸. It was recently shown that noradrenaline, dopamine and γ -aminobutyric acid (GABA) inhibit the voltage-dependent calcium channels in dorsal root and sympathetic ganglia⁹⁻¹¹; this inhibition is mimicked by intracellular application of guanine nucleotides and blocked by PT, suggesting the involvement of a G-protein^{12,13}. Here we report an inhibitory effect of the opioid D-Ala², D-Leu⁵-enkephalin (DADLE) on the calcium current (I_{Ca}) in neuroblastoma × glioma hybrid cells (N × G cells¹⁴). Pretreatment with PT almost completely abolishes the DADLE effect. The effect is restored by intracellular application of G_i and G_o . As the α -subunit of G_o (with or without β - γ complex) is 10 times more potent than G_i , we propose that G_o is involved in the functional coupling of opiate receptors to neuronal voltage-dependent calcium channels.

I_{Ca} was determined in differentiated N × G cells by the whole-cell clamp technique¹⁵, using Ba^{2+} as charge carrier. Outward currents were minimized by replacing K^+ with Cs^+ . On step-depolarizations from -40 to 0 mV, I_{Ca} had a density of $11.5 \pm 2.4 \mu A cm^{-2}$ (mean $\pm$ s.d.; $n = 46$); the current-voltage curve showed an apparent threshold at -40 mV, a maximum at 0 mV and an apparent reversal potential at 50 mV (Fig. 1). Extracellular application of $1 \mu M$ DADLE rapidly reduced I_{Ca} density to $3.3 \pm 2.5 \mu A cm^{-2}$ ($n = 8$) but did not affect threshold, reversal or maximum potential. I_{Ca} inhibition by DADLE was reversed by removing the peptide from the medium. In cells pretreated with PT ($200 ng ml^{-1}$) for at least 5 hours, the inhibitory DADLE effect on I_{Ca} was marginal (I_{Ca} density in the absence and presence of $1 \mu M$ DADLE was 10.6 ± 1.3 and $9.8 \pm 1.2 \mu A cm^{-2}$, respectively; $n = 6$). Intracellular application of the stable GTP analogue, GTP γ S ($100 \mu M$), mimicked¹⁶ the inhibitory effect on I_{Ca} but the stable GDP analogue, GDP β S, prevented it¹⁶ (as observed in dorsal root ganglia¹³). These findings strongly suggest the involvement of a PT-sensitive G-protein in the regulation of neuronal calcium channels.

We tested this hypothesis more directly by intracellularly applying the G-proteins. A mixture of G_i and G_o was purified from pig brain without the use of activating ligands¹⁷ and applied through the patch pipette into PT-treated cells at a concentration of $15 nM$ each. A small inhibitory effect of DADLE on I_{Ca} was observed after 5 min, and after 20 to 30 min, DADLE was as effective as in cells which had not been treated with PT (Fig. 2). The DADLE effect was fully reversed by washing.

To determine which of the two PT-sensitive G-proteins is involved in the hormonal control of neuronal calcium channels, individual G-proteins or G-protein subunits¹⁷ were applied to PT-treated cells (Fig. 3). Infusion of $3.8 nM$ G_i for 15 min caused partial restoration of the DADLE effect (I_{Ca} density in the absence and presence of $1 \mu M$ DADLE was 8.9 ± 0.7 and $7.2 \pm$

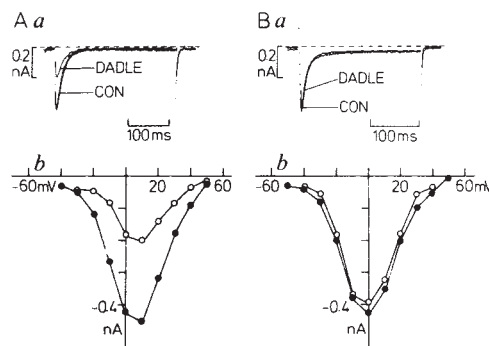


Fig. 1 Effect of DADLE on I_{Ca} in A, control and B, PT-treated N × G cells. Ba and Ba show superimposed current traces (test potential 0 mV). Control currents (CON) and currents after removing DADLE from the medium were superimposable. Currents recorded from cells differing slightly in their time course of inactivation. Ab and Bb show the corresponding current-voltage curves. ●, Control currents; ○, currents in presence of $1 \mu M$ DADLE.

Methods. N × G cells (passages 13–16) were grown as monolayers in Dulbecco's modified Eagle medium with 10% fetal calf serum, $0.1 mM$ hypoxanthine, $1 \mu M$ aminopterin and $16 \mu M$ thymidine at 7% CO_2 in air. For differentiation, cells were seeded at 10^3 cells mm^{-2} in 24-well Costar culture cluster plates, each containing several $3 \times 6 mm$ glass slides. Differentiation was induced after one day by serum withdrawal and addition of $1 mM$ dibutyl cyclic AMP. Cells were completely differentiated after 3–4 days; differentiated cells were treated with PT ($200 ng ml^{-1}$) for at least 5 h before the experiment. For electrophysiological studies, a glass slide with cells attached was transferred into a chamber ($200 \mu l$) and superfused at a rate of about $5 ml min^{-1}$. The extracellular solution contained (in mM) NaCl 140, CsCl 5.4, $BaCl_2$ 10.8, $MgCl_2$ 1, glucose 10, HEPES 10, pH 7.4, at $36^\circ C$. Whole cell recordings were performed with patch electrodes (2 – $5 M\Omega$) filled with (in mM) CsCl 112, EGTA 10, $MgSO_4$ 1, ATP 3, HEPES 5, pH 7.4. Intracellular application of GTP, the natural ligand of G-proteins, did not modify responses. From a computer simulation, using a simplified compartment model (unpublished results and ref. 31), we estimated that the time necessary to decrease the GTP-concentration in cytoplasm to 1% of the initial value (from 100 to $1 \mu M$) was 60 – 120 min. Thus the estimated cytoplasmic GTP concentration under our experimental conditions was above the reported K_D values for purified G-proteins. These values are likely to be much higher than the K_D values for G-proteins 'in situ'^{32,33}. When G-proteins were infused into the cytoplasm, the pipette solution also contained 0.5% bovine serum albumin (BSA) (added as carrier protein for the dilution of G-protein-containing stock solutions), $\leq 10 \mu M$ EDTA, $\leq 0.2 mM$ 2-mercaptoethanol, $\leq 0.3\%$ (v/v) ethyleneglycol, $\leq 1.5 mM$ NaCl and detergent (Lubrol PX or cholate) at the concentrations given. None of these additional constituents had an effect on I_{Ca} or inhibition of I_{Ca} by DADLE. Cells were clamped at a holding potential of -80 mV. Fast Na^+ currents were completely blocked by $100 ms$ long step depolarization from -80 mV to -40 mV. I_{Ca} was elicited by a subsequent $300 ms$ long step depolarization from -40 to 0 mV. Only for the current-voltage relations in Fig. 1 (lower panels), test potentials varied from -60 to $+60$ mV. The stimulation frequency was $0.2 Hz$. Calcium currents were measured as peak inward currents in reference to zero currents. A specific membrane capacitance of $1 \mu F cm^{-2}$ was assumed for the calculation of the current density. (For further details see refs 25, 31). The voltage-clamp protocol did not allow to differentiate between low and high voltage-activated calcium currents¹¹, and, therefore, the present current traces may reflect both components¹⁶.

$0.5 \mu A cm^{-2}$, respectively; $n = 4$). G_i at one tenth of this concentration failed to restore the response. In contrast, the α -subunit of G_o (α_o), which—like G_i and G_o —was purified without activating ligands¹⁷, was very effective even at a concentration of $0.4 nM$ (I_{Ca} density in the absence and presence of $1 \mu M$ DADLE was 9.3 ± 1.4 and $4.9 \pm 1.6 \mu A cm^{-2}$, respectively; $n = 5$). Addition of β - γ complex ($44 nM$) did not influence the effectiveness of α_o . The β - γ complex (without α -subunits) did

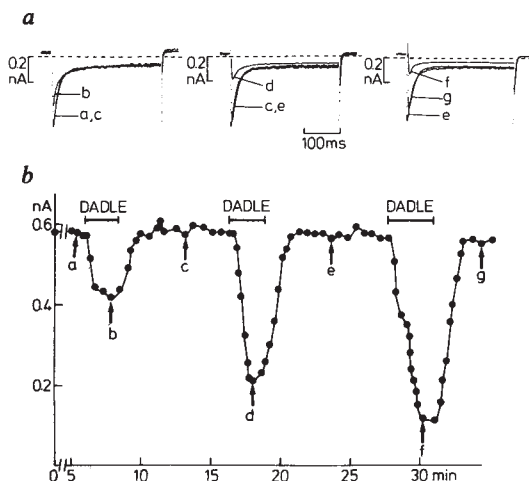


Fig. 2 Restoration of DADLE inhibition of I_{Ca} by intracellular application of G_i plus G_o . Current traces (a) were recorded at the time points a–g indicated in the time course of I_{Ca} (b); test potentials were 0 mV. In the illustrated experiment (one of five with similar results), a PT-pretreated cell was intracellularly infused with a mixture of G_i plus G_o and 1 μM DADLE was added at different times indicated by the bars. The effectiveness of intracellular application of macromolecules through the patch pipette was shown experimentally and by computer simulation using a simplified compartment model (unpublished results and ref. 31). G-proteins were purified¹⁷ from membranes of porcine brain to 90% homogeneity (SDS-PAGE; Coomassie blue stain) and stored in Lubrol PX-containing buffer at -80°C . The ratio of subunits was $\alpha_i : \alpha_o : \beta : \gamma = 1 : 1 : 2.5$. Purified G-proteins were diluted to 15 nM each in pipette buffer supplemented with 0.5% BSA. The final Lubrol PX concentration in the pipette buffer was 0.05%. The modified pipette buffer and heated G-protein solutions had no effect on I_{Ca} under control conditions or in the presence of DADLE.

not restore the DADLE effect at concentrations of up to 93 nM (I_{Ca} density in the absence and presence of 1 μM DADLE was 10.2 ± 1.6 and $9.4 \pm 1.5 \mu\text{A cm}^{-2}$, respectively; $n = 3$). Transducin (a G-protein purified from bovine retinae¹⁸), was also ineffective at concentrations of up to 195 nM in the pipette (I_{Ca} density in the absence and presence of 1 μM DADLE was 12.4 ± 1.3 and $11.7 \pm 1.4 \mu\text{A cm}^{-2}$, respectively; $n = 3$). Preparations of G-proteins denatured by heating for 10 min at 95°C did not alter basal I_{Ca} or the DADLE effect on I_{Ca} .

These findings imply that the voltage-dependent calcium channels of N×G cells are under effective neurotransmitter control; they suggest a physiological and pharmacological role for opiate receptor-mediated inhibition of I_{Ca} in the regulation of neurotransmitter release. Similar effects of adrenaline in the presence of a β -adrenergic blocking agent¹⁶ may provide a mechanistic explanation for inhibitory (for example antihypertensive) effects of centrally acting α_2 -adrenoceptor agonists.

G-proteins have been identified as membrane transducing components in receptor-mediated effects on enzymes acting as intracellular signal-generating effector systems¹⁹, for instance, adenylyl cyclase²⁰, cyclic GMP phosphodiesterase²¹ and phospholipase C²². Recent evidence suggests that G-proteins are also important in functional coupling of receptors to ion channels, for example, inhibition of voltage-dependent calcium channels in neurons, where it is known that cyclic AMP is not involved in channel regulation^{12,13}. Similarly, stimulation of K^+ channels in atrial cells by muscarinic agonists is independent of intracellular cAMP levels, yet appears to involve a PT-sensitive G-protein^{23,24}. In contrast, hormonal control of cardiac calcium channels is mediated by intracellular cAMP levels^{23,25,26}. Whereas in the case of adenylyl cyclase and cyclic GMP phosphodiesterase the G-proteins involved have been identified, it is at present not clear which G-proteins regulate ion channels.

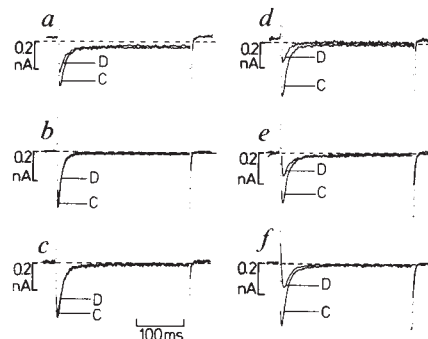


Fig. 3 Restoration of the DADLE effect on I_{Ca} by various G-protein preparations. Currents were recorded 15–20 min after disruption of the membrane patch. Shown are superimposed traces of I_{Ca} recorded in PT-treated N×G cells (test potential 0 mV). C, control current; D, current in the presence of 1 μM DADLE. Pipette solutions: a, b, 3.8 and 0.38 nM G_i , respectively (70% purity) in 0.015 and 0.0015% Lubrol PX with a 10-fold excess of β - γ complex; c, 93 nM β - γ complex (95% purity) in 0.05% Lubrol PX; d, e, 4.3 and 0.43 nM α_o respectively (90% purity) in 0.02 and 0.002% cholate; f, 4.3 nM α_o plus 44 nM β - γ complex in 0.05% Lubrol PX. In all experiments the pipette solution also contained 0.5% BSA. Restored DADLE effects were completely reversed by washing off the agonist. The slight differences in the holding current (a, d as compared to b, c, e, f) did not affect the reconstituted DADLE effect and were perhaps due to small changes in the pipette potential.

In cardiac tissues, muscarinic agonists inhibit adenylyl cyclase²⁷, presumably via G_i , and increase potassium currents (I_K) by a cyclic AMP-independent, PT-sensitive mechanism^{23,24}. In the central nervous system^{17,28}, as well as in N×G cells²⁹, G_o is a highly abundant membrane protein, which occurs in several-fold excess over G_i . In N×G cells carbachol inhibits adenylyl cyclase³⁰, as it does in cardiac cells, and increases I_K , but has no effect on I_{Ca} ¹⁶. The failure of carbachol to inhibit I_{Ca} despite its ability to inhibit adenylyl cyclase in N×G cells, together with the high potency of α_o compared to G_i in restoring the DADLE effect, support the possibility that G_o may be one of the G-proteins physiologically involved in functional coupling of various receptors to neuronal calcium channels. It is tempting to speculate that G_i may act as transducer molecule between muscarinic cholinergic agonists and K^+ channels in the heart, where G_o is much less abundant than in the central nervous system.

We thank Marianne Wulfern, Susanne Brendel, Ming Tang and Danielle Steimer for valuable assistance in these studies, Drs Motoyuki Yajima, Bernd Hamprecht and Klaus Peper for providing PT, N×G cells and tissue culture facilities, respectively, and Rosemarie Krüger for help in typing the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft, SFB 246 and the Fonds der Chemischen Industrie.

Received 14 August; accepted 30 October 1986.

1. Starke, K. *Rev. Physiol. Biochem. Pharmac.* **77**, 1–124 (1977).
2. Sharma, S. K., Klee, W. A. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3365–3369 (1977).
3. Codina, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4276–4280 (1983).
4. Katada, T., Bokoch, G. M., Northup, J. K., Ui, M. & Gilman, A. G. *J. biol. Chem.* **259**, 3568–3577 (1984).
5. Sternweis, P. C. & Robishaw, J. D. *J. biol. Chem.* **259**, 13806–13813 (1984).
6. Neer, E. J., Lok, J. M. & Wolf, L. G. *J. biol. Chem.* **259**, 14222–14229 (1984).
7. Hildebrandt, J. D., Codina, J., Ringer, R. & Birnbaumer, L. *J. biol. Chem.* **259**, 2039–2042 (1984).
8. Ui, M. *et al. Adv. cycl. Nucl. Prot. Phosphor. Res.* **17**, 145–151 (1984).
9. Dunlap, K. & Fischbach, G. D. *J. Physiol., Lond.* **317**, 519–535 (1981).
10. Deisz, R. A. & Lux, H. D. *Neurosci. Lett.* **56**, 205–210 (1985).
11. Marchetti, C., Carbone, E. & Lux, H. D. *Pflügers Arch. Eur. J. Physiol.* **406**, 104–111 (1986).
12. Scott, R. H. & Dolphin, A. C. *Neurosci. Lett.* **69**, 59–64 (1986).
13. Holz, I. V., G. G., Rane, S. G. & Dunlap, K. *Nature* **319**, 670–672 (1986).
14. Hamprecht, B., Glaser, T., Reiser, G., Bayer, E. & Probst, F. *Meth. Enzym.* **109**, 316–341 (1985).
15. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. Eur. J. Physiol.* **391**, 85–100 (1981).

16. Heschler, J. *et al. Adv. cyclic Nucleotide Prot. Phosphor. Res.* **21** (in the press).
17. Rosenthal, W. *et al. Eur. J. Biochem.* **158**, 255–263 (1985).
18. Kühn, H. *Nature* **283**, 587–589 (1980).
19. Bourne, H. R. *Nature* **321**, 814–816 (1986).
20. Gilman, A. G. *Cell* **36**, 577–579 (1984).
21. Kaupp, U. B. & Koch, K.-W. *Trends biochem. Sci.* **11**, 43–47 (1986).
22. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315–321 (1984).
23. Breitwieser, G. E. & Szabo, G. *Nature* **316**, 538–540 (1985).
24. Pfaffinger, P. J., Martin, J. M., Hunter, D. D., Nathanson, N. M. & Hille, B. *Nature* **317**, 536–538 (1985).
25. Heschler, J., Kameyama, M. & Trautwein, W. *Pflügers Arch. Eur. J. Physiol.* **407**, 182–189 (1986).
26. Reuter, H. *Nature* **301**, 569–574 (1983).
27. Jakobs, K. H., Aktories, K. & Schultz, G. *Naunyn-Schmiedeberg's Archs Pharmac.* **310**, 113–119 (1979).
28. Huff, R. M., Axton, J. M. & Neer, E. J. *J. biol. Chem.* **260**, 10864–10871 (1985).
29. Milligan, G., Gierschik, P., Spiegel, A. M. & Klee, W. A. *FEBS Lett.* **195**, 225–230 (1986).
30. Kurose, H., Katada, T., Amano, T. & Ui, M. *J. biol. Chem.* **258**, 4870–4875 (1983).
31. Kameyama, M., Heschler, J., Hofmann, F. & Trautwein, W. *Pflügers Arch. Eur. J. Physiol.* **407**, 123–128 (1986).
32. Northup, J. K., Smigel, M. D. & Gilman, A. G. *J. biol. Chem.* **257**, 11416–11423 (1982).
33. Bokoch, G., Katada, T., Northup, J. K., Ui, M. & Gilman, A. G. *J. biol. Chem.* **259**, 3560–3567 (1984).

Digital image processing of intracellular pH_i in gastric oxyntic and chief cells

Anthony M. Paradiso, Roger Y. Tsien & Terry E. Machen

Department of Physiology-Anatomy, University of California, Berkeley, California 94720, USA

Cytosolic pH_i (pH_i) is a critically regulated determinant of intracellular function. Several mechanisms for pH_i regulation in different tissues have been found, such as direct proton pumping^{1,2}, Na/H exchange³, Cl/HCO₃ exchange⁴, NaHCO₃ cotransport⁵, and Na/H/Cl/HCO₃ obligatorily linked^{6,7}. All these studies have used either single cells or cell populations assumed to be behaving homogeneously. Most tissues consist of more than one cell type, so it would be desirable to examine pH_i regulation simultaneously in many identified individual cells, particularly in epithelia where disaggregation and purification of isolated cells destroys the normal distinction between luminal and serosal environments. We have used a pH -sensitive fluorescent dye, BCECF (2',7'-bis(carboxyethyl)-5(6)-carboxyfluorescein) and digital image processing to study pH_i regulation simultaneously in the oxyntic cells (OC) and chief cells (CC) of gastric glands isolated from rabbit stomach. CCs become markedly more acidic upon removal of external Na (Na_o), but pH_i is restored rapidly on return to normal Na_o , with or without Cl. Oxyntic cell pH_i is much less affected by Na_o . Conversely, OCs become strongly more alkaline on removal of external Cl (Cl_o), pH_i being restored when Cl_o is replaced with or without Na, whereas CCs are relatively insensitive to Cl_o . Therefore, Na/H exchange is dominant over Cl/HCO₃ exchange in CCs, but in the neighbouring OCs, Cl/HCO₃ outweighs the Na/H mechanism, a heterogeneity that correlates with the functions of the two cell types.

To date, studies using ratio imaging of intracellular fluorescent dyes have focused mainly on measurements of cytosolic free calcium^{8–11}, though there have been a few studies of pH_i ^{12–15}. In all cases the measurements have been performed on free-living single cells. Recent studies concerned with electrolyte transport in OCs have shown that they contain both Na/H and Cl/HCO₃ exchangers^{16,17}. The cation exchanger is inhibited by amiloride, and the anion exchanger is inhibited by DIDS (4,4'-diisothiocyanostilbene-2,2'-disulphonic acid). Studies on isolated, purified membrane vesicles have shown that the exchangers are present in the basolateral membranes of OCs¹⁷. Nothing is known about pH_i regulation in CCs. It might be expected that, as these cells are exposed daily to highly acidic (pH 0.8) HCl solutions on their luminal membranes, they would have well-developed mechanisms for preventing cellular acidification.

BCECF ($pK_a \sim 7$) is highly sensitive, can be loaded into all the cells of a population without membrane disruption, yet leaks out relatively slowly^{18–20}. Increasing pH markedly enhances the excitation efficiency at 490 nm but has no effect at 439 nm (refs 16, 21), so the ratio of intensities obtained at 490 and 439 nm excitation measures pH_i , while cancelling variations in cell thickness and dye content. Glands in which the two cell types could be distinguished by transmitted light inspection (Fig. 1A) were imaged by epifluorescence excitation at 439 nm (Fig. 1B) and 490 nm (Fig. 1C) with the emission bandpass fixed at 520–550 nm. These images were digitized, and background intensities obtained from dye-free regions of the chamber were subtracted. The logarithm of the ratio of the two corrected images was calculated for each of the 512×486 picture elements and displayed as one of 32 possible pseudo-colour hues (Fig. 1D–F). The mean of the intensities at the two wavelengths controlled the brightness of the pseudo-colours over 8 steps so that image areas outside the gland appear black, and areas of gland with less dye are dimmer than parts from which more signal was received. To calibrate the 490/439 nm ratio with pH_i , the latter was clamped to three widely separated values, for example pH 8.47, 6.95 and 5.47 in Fig. 1D–F. Setting pH_i with high $[K]_o$ and the H/K exchanger nigericin gave quantitatively similar results as using the Na/H ionophore monensin in glands treated with ouabain to raise internal Na concentration $[Na]_i$. From the three ratios obtained at known pH_i values, other ratios can be converted to pH_i using the standard equations for indicators with 1:1 stoichiometry. Such *in situ* calibration is based on the actual properties of the indicator in intact cells without assuming that those are the same as *in vitro* buffers. The uniformity of ratio within each image (Fig. 1D–F) shows that the characteristics of the dye are the same in the two cell types.

A series of experiments was designed to compare the responses of OCs and CCs to changes of $[Na]_o$ and $[Cl]_o$, and typical responses of the two cell types are shown in Fig. 2A–F. When the gland was bathed with NaCl Ringer's solution, OCs and CCs of this gland had nearly identical pH_i (Fig. 2B). However, pH_i was often more alkaline in CCs than in OCs (40% of the glands), and this difference is reflected in the average pH_i values for CCs and OCs shown in Table 1, series 1. Average pH_i for a given OC or CC was calculated by averaging the intensity-weighted ratio values of all pixels within a circular spot which corresponded to a region of the cells that excluded as much as possible adjacent (or underlying) cells. The optical contamination between cell types was more severe with CCs, which are wedged between OCs.

Changing the transmembrane gradients of $[Na]$ or $[Cl]$ (or both) caused pH_i of OCs and CCs to behave differently. The apical membranes of gland cells face a collapsed glandular lumen and it is therefore likely that changes of pH_i in response to changes in the composition of the bathing solutions probably occur by mechanisms at the basolateral membranes of the cells. A minor contribution from the apical membranes cannot, though, be excluded. Na-free (*N*-methyl-D-glucamine, NMG) solutions caused both OCs and CCs to become more acidic, but the effect was much larger for CCs than OCs (Fig. 2C; Table 1, series 2). When NaCl Ringer's was changed to a Cl-free (gluconate) solution, both cell types became more alkaline, and the effect was larger for OCs than CCs (Fig. 2D; Table 1, series 3). When the NaCl Ringer's was replaced with a solution in which both Na and Cl were removed (NMG gluconate), pH_i in OCs and CCs changed in opposite directions—OCs became more alkaline but CCs became more acidic (Fig. 2E). Figure 2F shows the steady-state response for OCs and CCs in NaCl Ringer's after NMG chloride (NMGCl), sodium gluconate (Nagluconate) and NMGgluconate treatments.

The time course of the changes in pH_i in five OCs and five CCs of one gland in response to NMGgluconate is shown in Fig. 3A: OCs exhibited a biphasic increase-then-decrease in pH_i , but CCs showed only a decrease in pH_i . Steady-state pH_i for

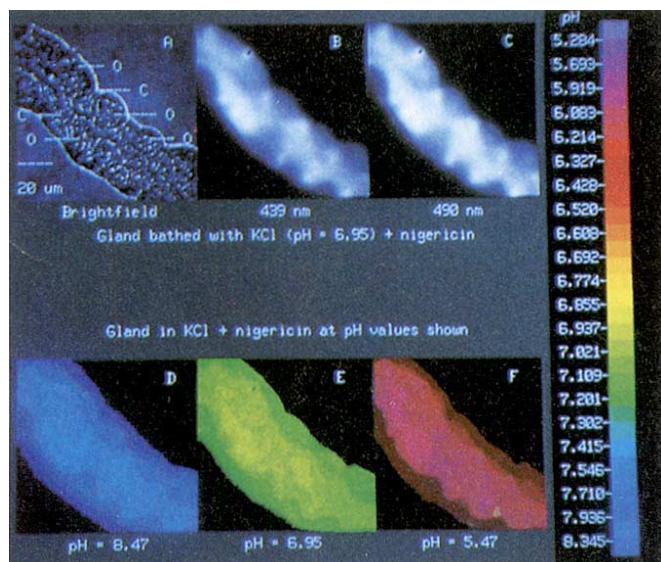


Fig. 1 A, Brightfield micrograph showing OCs and CCs of a BCECF-loaded gastric gland; bar 20 μm . B, C, Fluorescent micrographs of cytoplasmic-trapped BCECF at excitation wavelength 439 nm (A) and 490 nm (B); emission, 520–550 nm. D–F, Processed video images of the gland bathed with KCl Ringer's + 10^{-5} M nigericin at pH_o values of 8.47 (D), 6.95 (E), and 5.47 (F). All images were acquired at steady state. The bar on the right shows the pH_i scale in pseudo-colour. In high [K]-nigericin treated glands there was no statistical difference between the ratios of OCs and CCs at any given pH_i . The small black dot on the upper edge of the gland was a defect on the video camera photocathode.

Methods. Glands from white New Zealand rabbits were isolated and loaded with the acetoxymethyl ester of BCECF (Molecular Probes) as reported previously^{16,18,25}. BCECF-loaded glands were attached to a polylysine-coated coverslip and mounted in a chamber which allowed constant and rapid perfusion (turnover, 10 times min^{-1}). The chamber was placed over the objective ($\times 63$ oil, 1.25 NA) of a Zeiss IM35 microscope. The fluorimeter system used to generate two synchronized light beams of selected wavelengths has been described²⁶. Video images of BCECF fluorescence at 490 and 439 nm were acquired by a SIT camera and relayed both into a TV monitor and a Gould FD5000 image processor. All operations of the FD5000 were controlled by a DEC MicroPDP 11/73 computer. Log ratio (490/439 nm) was converted to pH_i using the equation $pH_i = [(R_x - R_\infty)/(R_0 - R_\infty)] + pK'_a$, where R_x is the experimentally-derived ratio, R_∞ and R_0 are the limiting ratios for the unbound and bound forms of the dye, respectively, and pK'_a is the negative log of the effective dissociation constant (K'_a) of the dye²⁷. NaCl Ringer's contained (in mM): 150 NaCl, 1.0 CaSO_4 , 1.0 MgSO_4 , 3 K_2HPO_4 , 10 glucose and 20 HEPES. In KCl Ringer's, all Na was replaced with K. In Cl-free Ringer's gluconate replaced Cl. In Na-free Ringer's N-methyl-D-glucamine (NMG) replaced Na. In Na- and Cl-free Ringer's NMGgluconate replaced NaCl. Solutions were equilibrated with air and adjusted to pH 7.40–7.50 with NMG base. Experiments were conducted at room temperature (24 $^\circ\text{C}$).

OCs and CCs in response to NMGgluconate treatment are given in Table 1, series 4. The rest of Fig. 3a shows that when NMGgluconate was replaced with NMGCl so that $[\text{Cl}]_o > [\text{Cl}]_i$ (with no Na in the cells or in the solution), both OCs and CCs became more acidic, as expected if Cl were entering and HCO_3^- were leaving the cell (see also Table 1, series 5). Changing from NMGgluconate to Nagluconate (thereby creating $[\text{Na}]_o > [\text{Na}]_i$ in a Cl-free solution) caused both cell types to become more alkaline, as expected if Na were entering and H were leaving the cell (see also Table 1, series 6). Interestingly, the Cl-dependent changes of pH_i were larger for OCs than for CCs but the Na-dependent changes of pH_i were larger for CCs than for OCs. It has been previously shown for OCs that the Na-dependent

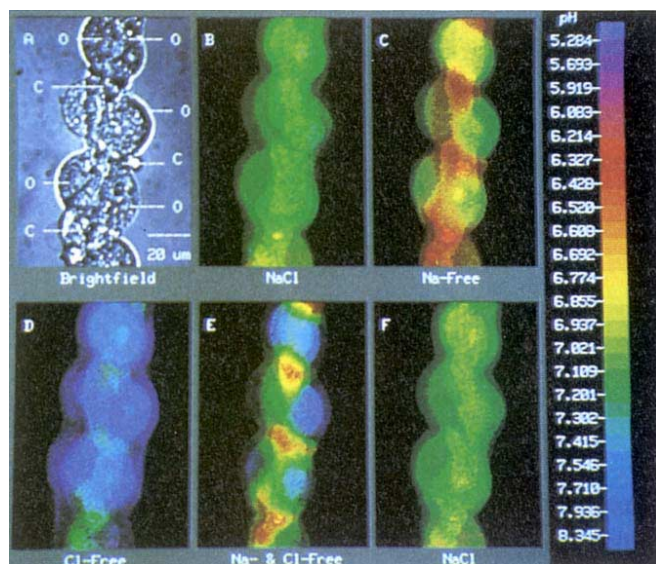


Fig. 2 A, Brightfield micrograph showing OCs and CCs of a BCECF-loaded gland. B, C, D, F, steady state pseudo-colour images of the gland bathed with NaCl (B and F), Na-free (C) and Cl-free (D) solutions. E, Pseudo-colour image acquired 2.6 min after changing from NaCl to NMGgluconate Ringer's.

alkalinization was blocked by 1 mM amiloride and the Cl-dependent changes of pH_i were blocked by 0.2 mM DIDS^{16,17}.

The differences between OCs and CCs in response to changing $[\text{Na}]_o$ and/or $[\text{Cl}]_o$ could be explained by two different possibilities. (1) Both cell types have both Na/H and Cl/ HCO_3^- exchangers, but the activity of the anion exchanger is high relative to that of the cation exchanger in OCs while the opposite is true for CCs. (2) Both cell types have equally active Na/H and Cl/ HCO_3^- exchangers, but the CCs have a parallel pathway for Cl movement which does not contribute to changes of pH_i (for example, a Cl conductance) and the OCs have a parallel pathway for Na movement. These hypotheses were tested by measuring rates of change of pH_i in response to some of the changes of solution composition used above. If hypothesis (1) is correct, the Na-dependent changes of pH_i should occur faster in CCs while the Cl-dependent changes of pH_i should occur faster in OCs. Under hypothesis (2), the two cell types would not differ in their initial rates of change of pH_i , though CC responses to $[\text{Cl}]_o$ and OC responses to $[\text{Na}]_o$ would subsequently be truncated by the postulated Cl and Na shunts. Rapid temporal resolution was necessary for these experiments so a microspectrofluorimetric technique described previously¹⁶ was used. Briefly, pH_i was measured by focusing on single OCs or CCs in intact glands, and 490/439 nm intensity ratio was acquired using a photomultiplier tube (instead of the TV camera which was used for the image processing experiments). The advantage of using the photomultiplier was that the temporal resolution increased to 300 ms (compared with 6 s for the image processor). Figure 3b and e shows typical experiments in which rates of change of pH_i were measured in single OCs and single CCs in response to replacing NaCl with Na-free (Fig. 3b and d) or Cl-free (Fig. 3c and e) solutions.

Initial rates of change of pH_i in OCs and CCs during changes of only $[\text{Na}]$ or $[\text{Cl}]$ are also summarized in Table 1. For OCs, changing between Na-free and Na-containing solutions (from NaCl to NMGCl or NMGgluconate to Nagluconate) caused pH_i to change at rates which were 2.4–7.5 times slower than changing between Cl-free and Cl-containing solutions (NaCl to Nagluconate or NMGgluconate to NMGCl). The opposite was true for CCs: changing between Na-free and Na-containing

Fig. 3 Effects of altering [Na] and [Cl] on pH_i of OCs and CCs of gastric glands. Boxes show solutions used, abbreviations as in text. *a*, Average pH_i of OCs (○) and CCs (●) in a single gastric gland as measured with the image processor. The experiment begins with the gland bathed with NaCl Ringer's. The solution was subsequently changed as shown. Each data point represents an average pH_i value from 5 OCs or 5 CCs. The s.e.m. of each data point was ≤ 0.04 pH units. *b-e*, Effects of altering solution [Na] and [Cl] on pH_i of single OCs (*b* and *c*) and CCs (*d* and *e*) in gastric glands as measured with the photomultiplier. NaCl Ringer's solution was changed to NMGCl or Nagluconate as shown.

Methods. As described¹⁶, the fluorescence from a single OC or CC was acquired alternately at 490 and 439 nm by focusing (spot diameter 5–15 μ m) onto the central region of the cells. The photomultiplier output was passed to a microcomputer which stored the signals resulting from the two excitation monochromators. After subtracting background (measured on unloaded OCs and CCs from the same batch of glands), the ratio ($R(490/439)$) was calculated. Calibration of $R(490/439)$ in terms of pH_i was performed by exposing the glands at the end of an experiment to KCl Ringer's containing 10^{-5} M nigericin at three different pH values. There were no statistical differences between pH_i measured using the photomultiplier and the image processing system for any of the experimental procedures reported here. Initial rates of change of pH_i in response to changes of solution [Na] or [Cl] were calculated as follows. The time course of pH_i (or R) following a change of [Na] or [Cl] in the solution could be described by the equation $R_t = R_{\max} - (R_{\max} - R_{\text{initial}}) \exp(-kt)$ where R_{initial} , R_t and $R_{\max}$ refer to ratios at time zero, a given time t and in the steady state, respectively. The rate coefficient, k , was determined from the slope of the linear regression plot of $\ln [(R_t - R_{\max}) / (R_t - R_{\text{initial}})]$ against t . Initial rates (pH units min^{-1}) were then calculated as $k(pH_{\max} - pH_{\text{initial}})$. These rates were not significantly different from those measured from the initial slopes of pH_i against time.

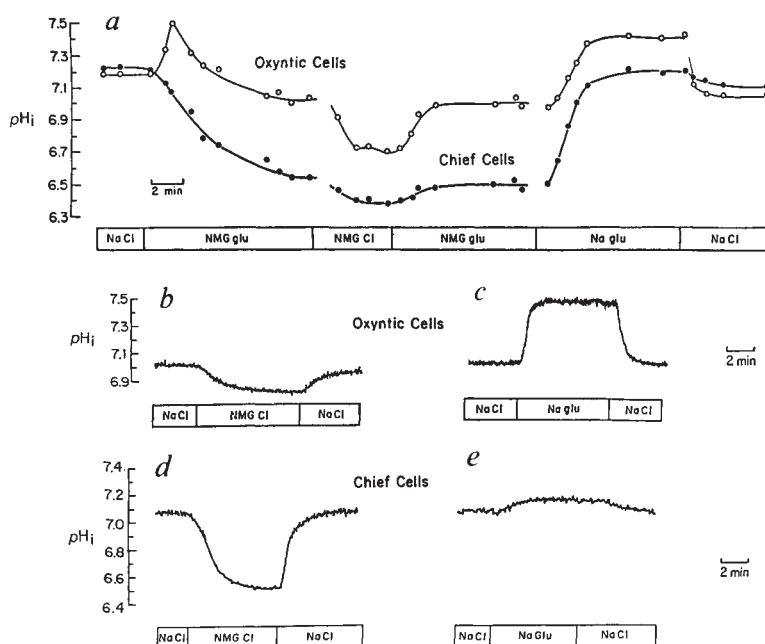


Table 1 Rates of change of pH_i , with final values, for oxyntic and chief cells in various solutions

Series	Treatment	Oxyntic cells		Chief cells	
		ΔpH_i (min^{-1})	Final pH_i	ΔpH_i (min^{-1})	Final pH_i
1	NaCl	—	7.16 ± 0.01 (60/5)	—	7.23 ± 0.02 (60/5)
2	NaCl $\rightarrow$ NMGCl	0.17 ± 0.02 (10/3)	6.85 ± 0.02 (21/3)	0.44 ± 0.03 (10/3)	6.52 ± 0.02 (21/3)
3	NaCl $\rightarrow$ Naglu	1.68 ± 0.14 (15/4)	7.63 ± 0.04 (21/3)	0.08 ± 0.02 (7/2)	7.30 ± 0.03 (21/3)
4	NaCl $\rightarrow$ NMGglu	—	7.12 ± 0.03 (14/2)	—	6.62 ± 0.02 (16/2)
5	NMGglu $\rightarrow$ NMGCl	0.74 ± 0.03 (12/3)	6.77 ± 0.02 (13/2)	0.02 ± 0.01 (3/2)	6.49 ± 0.0 (12/2)
6	NMGglu $\rightarrow$ Naglu	0.28 ± 0.03 (10/3)	7.52 ± 0.03 (15/2)	0.50 ± 0.04 (10/3)	7.23 ± 0.02 (14/2)

Steady-state pH_i in OCs and CCs in NaCl Ringer's are given in series 1. For series 2–6 the initial rate of change of pH_i and the final pH_i are shown for the transition from the first solution to the second solution connected by the arrow: Naglu and NMGglu are Nagluconate and NMGgluconate respectively; NMGCl as before. For series 2–4 the NaCl Ringer's was changed to Na-free (series 2), Cl-free (series 3) or Na and Cl-free (series 4) solutions. For series 5 and 6, the glands were initially bathed in NaCl Ringer's before changing to NMGgluconate Ringer's. After a new steady state pH_i was reached, NMGgluconate was changed to NMGCl (series 5) or Naglu (series 6). For any given Ringer's solution, steady state pH_i of OCs was significantly different from that of CCs ($P < 0.05$, two-tailed t -test in all cases). Numbers in parentheses represent number of cells/number of different animals.

solutions caused pH_i changes 12.5–16.8 times faster than changing between Cl-free and Cl-containing solutions. Also, it should be noted that changing from NaCl to Nagluconate caused the pH_i of OCs to increase approximately to the same extent and at the same rate in solutions which were nominally HCO_3^- -free (pH_o 7.4–7.5, equilibrated with air, $[\text{HCO}_3^-] = 0.2$ mM) and in solutions containing 25 mM HCO_3^- (pH_o 7.4, 5% CO_2 gas)¹⁶.

These data are consistent with the idea that both cell types have both Cl/HCO_3^- and Na/H exchangers, but the anion exchanger is dominant in the OCs and the cation exchanger in the CCs. Thus, for OCs the Cl-dependent changes of pH_i were larger and occurred more rapidly than the Na-dependent changes. And for CCs, the Na-dependent changes of pH_i were larger and occurred more rapidly than the Cl-dependent

changes. It is even possible that the CCs have no anion exchange activity, and their small apparent Cl-dependent changes of pH_i are due to optical contamination from adjacent OCs. The cation exchanger normally tends to alkalinize and the anion exchanger tends to acidify, so the apparently asymmetric distribution of the exchangers in the two cell types may explain the fact that the CCs are slightly more alkaline than the OCs in NaCl Ringer's (Table 1, series 1).

The asymmetric distribution of the exchangers in OCs and CCs is also consistent with the physiological functions of the two cells. As production of HCl at the apical membrane (by the H, K -ATPase and its associated conductances^{22,23}) requires the steady entry of Cl into and exit of HCO_3^- out of the cell across the basolateral membrane²⁴, the Cl/HCO_3^- exchanger could

provide both these activities. The Na/H exchanger may be particularly useful for regulating pH_i in CCs during acidic stress. For example, during a meal the pH of the solution bathing the apical membranes of CCs could be as low as pH 0.8 when pH_i is 7.2. Thus, there is a considerable gradient favouring H entry into these gastric cells. The basolateral Na/H exchanger would remove H leaking into the cells from the glandular lumen.

We thank Dr Martin Poenie for his help. This work was supported by grants from USPHS, NIH AM19520 and GM31004. A.M.P. was supported by an Individual National Service Award (NIH, DK07766).

Received 2 September; accepted 28 November 1986.

1. Steinmetz, P. A. & Anderson, O. S. *J. Memb. Biol.* **65**, 155-174 (1982).
2. Forte, J. G. & Machen, T. E. in *Physiology of Membrane Disorders* 2nd edn (eds Andreoli, T. E., Hoffman, J. F., Fanestil, D. D. & Schultz, S. G.) 535-558 (Plenum, New York, 1986).
3. Grinstein, S. & Rothstein, A. *J. Memb. Biol.* **90**, 1-12 (1986).
4. Cabantchik, Z. I., Knauf, P. A. & Rothstein, A. *Biochim. biophys. Acta* **515**, 239-302 (1978).
5. Boron, W. F. & Boulpaep, E. L. *J. gen. Physiol.* **81**, 53-94 (1983).

Failure to find holes in the T-cell repertoire

Kazumasa Ogasawara, W. Lee Maloy & Ronald H. Schwartz

Laboratories of Immunology and Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

The ability of an animal to respond to a given antigenic peptide depends on its major histocompatibility complex (MHC) type¹. Some peptides are not immunogenic when combined with a particular form of the MHC-encoded molecule. This non-responsiveness is regulated by immune response (*Ir*) genes and is thought to arise by one of two distinct mechanisms. Either the MHC-encoded molecules physically fail to interact with the antigen, preventing the activation of T cells with appropriate receptors, or they limit the expressed repertoire of T cell clones so that no T cells are available to be activated by existing complexes of MHC-encoded molecules and antigen². Experimental evidence has been generated to support both mechanisms (reviewed in ref. 1). However, the relative importance of each has not been clearly established. In this study we started with a peptide that was immunogenic in B10 mice; it was thus known to be able to interact with the MHC molecule, and T cells existed which could recognise the peptide-MHC complex. Based on previous experiments³, we then changed only those parts of the peptide that we thought interacted with the T-cell receptor. All the new analogues created were still immunogenic, confirming that the amino-acid substitutions that we had made did not prevent productive interactions with the MHC-encoded molecule. No limitations ('holes') in the T-cell repertoire were found. The experiments demonstrate the vast potential of the T-cell population to recognize many different analogues, each in a unique way, and suggest that constraints on the diversity of the T-cell repertoire may not be a major explanation for *Ir* gene defects.

Previous studies in our laboratory have dissected T-cell antigenic determinants into three components^{3,4}: residues that interact with the T-cell receptor; residues that interact with the MHC-encoded molecule; and residues that constrain the relevant structure of the determinant (usually a peptide of 10-15 amino acids). This classification was accomplished by making a series of single-amino-acid substitutions in the native peptide and using these analogues in several biological tests to determine effects on T-cell priming and stimulation. In particular, if the analogue was less potent than the original peptide when used to stimulate a T-cell clone that was specific for the original peptide, but was immunogenic and elicited different T-cell

6. Thomas, R. C. *J. Physiol., Lond.* **273**, 317-338 (1977).
7. Boron, W. F. & Russell, J. M. *J. gen. Physiol.* **81**, 373-399 (1983).
8. Keith, C. H., Maxfield, F. R. & Shelanski, M. L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 800-804 (1985).
9. Keith, C. H., Rajiv, R., Maxfield, F. R., Bajer, A. & Shelanski, M. L. *Nature* **316**, 348-350 (1985).
10. Peonie, M., Alderton, J., Steinhardt, R. & Tsien, R. *Science* **233**, 886-888 (1986).
11. Sawyer, D. W., Sullivan, J. A. & Mandell, G. L. *Science* **230**, 663-666 (1985).
12. Heiple, J. M. & Taylor, D. L. *J. Cell Biol.* **86**, 885-890 (1980).
13. McNeil, P. L., Tanasugarn, L., Meigs, J. B. & Taylor, D. L. *J. Cell Biol.* **97**, 692-702 (1983).
14. Tanasugarn, L., McNeil, P., Reynolds, G. T. & Taylor, D. L. *J. Cell Biol.* **98**, 717-724 (1984).
15. Slavik, J. *FEBS Lett.* **156**, 227-230 (1983).
16. Paradiso, A. M., Tsien, R. Y., Demarest, J. R. & Machen, T. E. *Am. J. Physiol.* (submitted).
17. Muallem, S., Burnham, C., Blissard, D., Berglinth, T. & Sachs, G. *J. biol. Chem.* **260**, 6641-6653 (1985).
18. Paradiso, A. M., Negulescu, P. A. & Machen, T. E. *Am. J. Physiol.* **250**, G524-G534 (1986).
19. Rink, T. J., Tsien, R. Y. & Pozzan, T. *J. Cell Biol.* **95**, 189-196 (1982).
20. Grinstein, S., Cohen, S. & Rothstein, A. *J. gen. Physiol.* **830**, 341-369 (1984).
21. Alpern, R. J. *J. gen. Physiol.* **84**, 613-636 (1985).
22. Wolosin, J. M. & Forte, J. G. *Am. J. Physiol.* **246**, C537-C547 (1984).
23. Cuppoletti, J. & Sachs, G. *J. biol. Chem.* **259**, 14952-14959 (1984).
24. Davies, R. E. *Biol. Rev.* **26**, 87-120 (1951).
25. Paradiso, A. M., Tsien, R. Y. & Machen, T. E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7436-7440 (1984).
26. Tsien, R. Y., Rink, T. J. & Poenie, M. *Cell Calcium* **6**, 145-157 (1985).
27. Gryniewicz, G., Poenie, M. & Tsien, R. Y. *J. biol. Chem.* **260**, 3440-3450 (1985).

clones specific for the new peptide, the altered residue was postulated to interact with the T-cell receptor. This is because the immune system appeared to be able to recognize the particular amino acid that was located at that position and to respond to it in a clonal manner. In contrast, substitutions at other amino-acid positions did not elicit new T-cell clones upon immunization, although the changes did affect the potency of the peptide and/or the interaction with the MHC-encoded molecule³.

Recently we described a peptide from the antigen pigeon cytochrome *c* which is immunogenic for T cells in B10 (*H-2^b*) mice⁵. The response to this antigenic determinant is under *Ir* gene control. B10 congenic strains bearing *H-2^b* or *H-2^d* are highly responsive to the peptide; strains bearing *H-2^{f,p,k,q}* or *v* are less responsive or do not respond at all (data not shown). In the process of mapping the antigenic determinant for residues that interact with T-cell receptors in the B10 strain, we decided to test how flexible the T-cell repertoire was at recognizing different substitutions at such positions. By using single-amino-acid substitutions it was possible to change the peptide without altering its ability to interact with the MHC-encoded molecule, thus providing a method for selectively probing for holes in the T-cell repertoire.

The antigenic determinant consisted of residues 43-58 of the pigeon cytochrome *c* sequence. When a peptide containing this sequence (AEGFSYTDANKNKGIT) was synthesized and used to immunize B10 mice, T cells from the draining lymph nodes were found to proliferate *in vitro* when stimulated with the peptide (Figs 1a and 2a). A large number of analogues, containing amino-acid substitutions at one of the 16 positions, were then used to stimulate the same T-cell population. The analogues with substitutions at position 50 or 52 revealed a significant decrease in potency even with conservative substitutions such as glutamic acid (E) for aspartic acid (D) at position 50 or glutamine (Q) for asparagine (N) at position 52 (Figs 1a and 2a). Similar substitutions at other residues, such as position 54, did not have this kind of effect on the T-cell response (data not shown). These preliminary observations suggested that amino acids 50D and 52N were good candidates for residues that interacted with the T-cell receptor³.

To test this hypothesis, B10 mice were immunized with the different analogues and their lymph node T cells challenged *in vitro* with each of the peptides. As shown in Figs 1b and 2b, the immunogen, substituted at either position 50 (N for D) or 52 (Q for N), respectively, was the most potent stimulator. The T-cell proliferative response to these analogues was detectable at lower antigen concentrations than with any of the other peptides, including the native sequence, and reached a higher maximum value. These results demonstrated that the B10 T-cell immune system could distinguish between substitutions at these

Fig. 1 Specificity of the T-cell proliferative response to the pigeon cytochrome *c* peptide 43–58 and its residue 50 analogues. Mice were immunized with: *a*, the synthetic peptide pigeon 43–58 with an aspartic acid at residue 50 (50D); *b*, pigeon 43–58 with an asparagine at position 50 (50N); *c*, pigeon 43–58 with a valine at position 50 (50V); *d*, pigeon 43–58 with a leucine at position 50 (50L). The response of the T cells to stimulation *in vitro* with different peptide analogues with changes at position 50 is shown. In addition to the four analogues used for immunization, the other peptides include pigeon 43–58 (50A), 43–58(50E), 43–58(50Q), 43–58(50K) and 43–58(50M), at position 50 an alanine, glutamic acid, glutamine, lysine or methionine residue, respectively.

Methods. B10 or B10.A(18R) mice were immunized in the footpads and base of the tail with 20 nmol of a synthetic peptide dissolved in saline and mixed 1:1 in an emulsion with complete Freund's adjuvant containing H37Ra mycobacteria (Difco). After 7–10 days, T cells from the draining lymph nodes were prepared by passing the cells over nylon wool columns¹². In 96-well flat-bottomed microtitre plates, 4×10^5 T cells per well were cultured with or without antigen and 1×10^5 irradiated (3000 rad) syngeneic spleen cells as a source of antigen-presenting cells in a total volume of 0.2 ml Click's medium supplemented with 10% heat-inactivated fetal calf serum, gentamycin, penicillin, glutamine and 2-mercaptoethanol. After 3 days of culture, 1 μ Ci of [³H]thymidine (6.7 Ci mmol⁻¹) was added to each well in 10 μ l of saline. The cells were collected 18 h later using a PHD cell harvester (Cambridge Technology, Inc.) and thymidine incorporation into DNA was measured by standard liquid scintillation counting techniques. Determinations were done in duplicate, and the data are presented as the mean incorporation with the background value (medium alone) subtracted (Δ c.p.m.). All the peptides were made using the Merrifield solid phase procedure as previously described⁴ and purified by reverse phase HPLC chromatography on a Vydac C18 column.

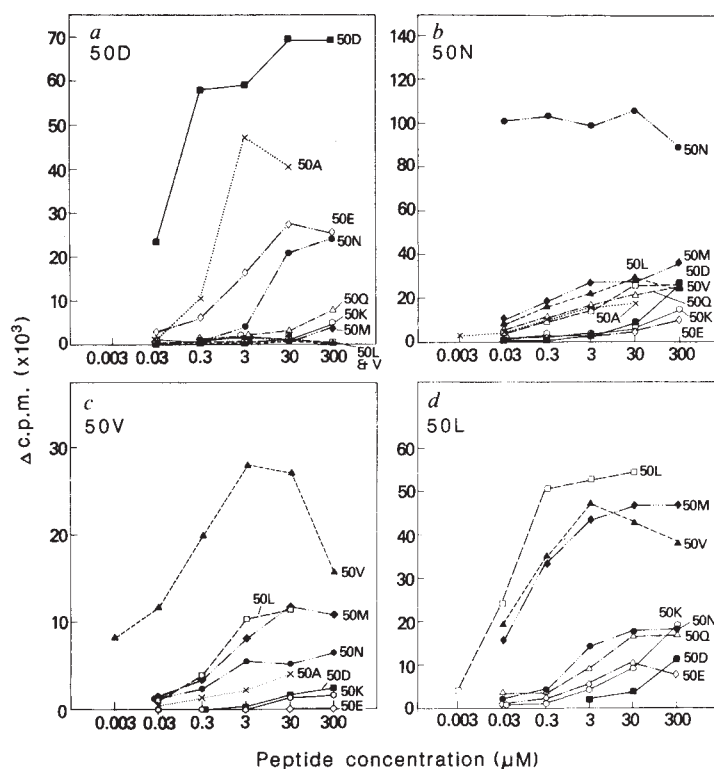


Table 1 T-cell proliferative response to various peptides

Immunogen	In vitro stimulant (% of maximum response)									
	43–58	50N	50E	50Q	50K	50M	50A	50V	50L	
43–58 (50D)	100	11	21	4	1	3	39	1	<1	
50N	9	100	6	22	4	29	14	25	14	
50E	19	14	100	39	<1	8	3	2	4	
50Q	3	12	22	100	15	29	9	14	6	
50K	4	12	7	17	100	77	9	18	21	
50M	4	29	9	43	13	100	29	41	35	
50A	9	15	12	21	2	27	100	32	8	
50V	6	23	2	16	3	35	9	100	33	
50L	5	16	7	16	6	63	4	67	100	
50N, 52E	2	20	2	6	1	5	2	5	<1	
50N, 52D	2	11	3	6	1	<1	<1	3	<1	
50K, 52E	1	<1	<1	1	14	1	<1	1	1	
50K, 52D	4	3	3	3	16	4	1	4	2	
	43–58	52D	52E	52Q	52T	52Y	50N, 52E	50N, 52D	50K, 52E	50K, 52D
43–58 (52N)	100	4	2	6	18	5	6	9	<1	<1
52D	12	100	4	4	9	3	NT	NT	NT	NT
52E	23	19	100	24	34	17	NT	NT	NT	NT
52Q	3	2	8	100	1	5	NT	NT	NT	NT
52T	12	3	6	11	100	6	NT	NT	NT	NT
52Y	7	3	5	9	15	100	NT	NT	NT	NT
50N, 52E	2	1	2	1	1	2	100	18	2	1
50N, 52D	2	<1	1	<1	<1	<1	9	100	<1	1
50K, 52E	1	<1	<1	<1	<1	<1	4	<1	100	25
50K, 52D	4	2	2	3	1	2	4	4	8	100

B10 or B10.A(18R) mice were immunized with 20 nanomoles of one peptide analogue and the T cells from the draining lymph nodes were stimulated *in vitro* with each of the various analogues as described in Figs 1 and 2. The greatest response is always seen with the immunizing peptide. To make this summary table of dose-response curves for analogue comparison purposes only, a rough measure of the area under each curve was made by summing the thymidine incorporations (Δ c.p.m.) observed at 0.03, 0.3, 3, 30 and 300 μ M peptide. This sum is expressed as a percentage of the largest value observed, which was always obtained with the immunogen. Eighty-five % of the analogues were tested in two independent experiments and the value in the table represents the mean. The values for the other 15% represent single determinations. For some peptides the lowest or highest concentration was not tested. In those cases percentages were derived by comparing sums calculated only over the range tested. The abbreviations for the analogues examined are defined in Figs 1 and 2. NT: not tested.

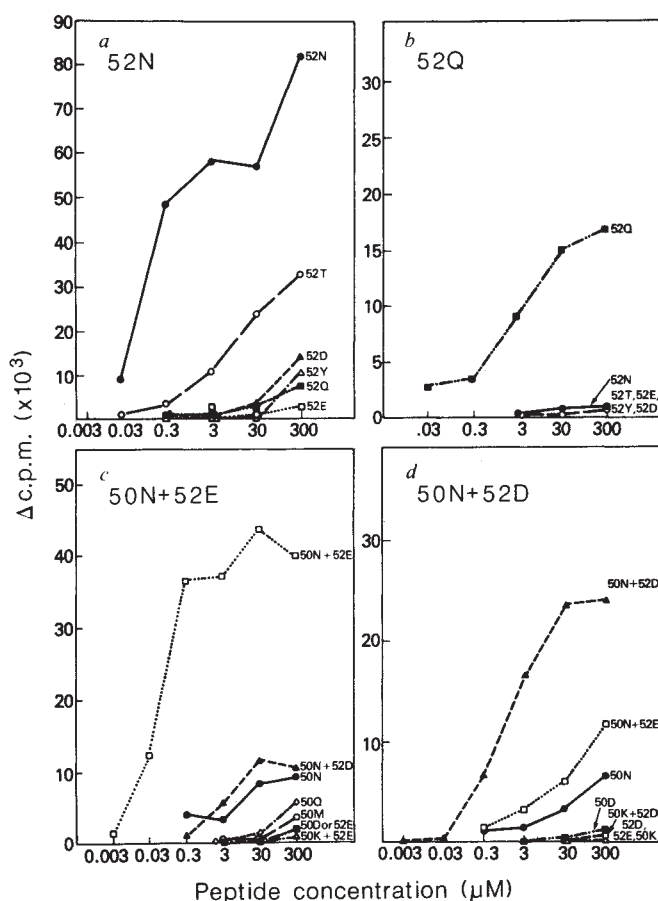


Fig. 2 Specificity of the T-cell proliferative response to the pigeon cytochrome *c* peptide 43–58 and its analogues with single-amino-acid substitutions at residue 52 and double substitutions at residues 50 and 52. The mice were immunized with the synthetic peptide pigeon 43–58, with *a*, an asparagine at residue 52 (52N); *b*, a glutamine at position 52 (52Q); *c*, an asparagine at position 50 and a glutamic acid at position 52 (50N+52E); *d*, an asparagine at position 50 and an aspartic acid at position 52 (50N+52D). The response of the T cells to stimulation *in vitro* with several of the 5 different analogues at position 52 or 8 different analogues at position 50 or the four double substitutions is shown. In addition to the four analogues used for immunization and the 8 substitutions at position 50 described in Fig. 1, these peptides include pigeon 43–58(52T), 43–58(52Y), 43–58(52E) and 43–58(52D) which contain at position 52 a threonine, tyrosine, glutamic acid or aspartic acid, respectively, and pigeon 43–58(50K+52E) and 43–58(50K+52D) which contain a lysine at position 50 and either a glutamic acid or an aspartic acid at position 52. For methods, see Fig. 1.

two positions and mount a response that was largely specific for the amino acid located there.

To explore the flexibility of the T-cell repertoire, multiple substitutions were made at each residue (50 or 52), and the peptide analogues tested for both immunogenicity and specificity of the evoked response. As shown in Table 1 and Figs 1 and 2, every analogue tested elicited a T-cell response that was specific for the immunogen. Even subtle differences such as valine (V) versus leucine (L) at position 50 were discriminated (Fig. 1*c, d*), although not all clones in the responding T-cell population could make this distinction when 50L was the immunogen, that is, significant cross-stimulation was observed with 50V. But the original immunogen (50L) still generated the largest response at lower antigen concentrations, suggesting that at least some clones could distinguish between these two analogues.

In all, 8 amino-acid substitutions were made at position 50 and 5 changes at position 52 (Table 1). In every case immunization yielded a T-cell population that was most responsive to

the immunogen. In addition, four analogues were constructed with double substitutions, one at 50 and one at 52. As shown in Fig. 2*c, d* and Table 1, these analogues also elicited T cells specific for the immunogen. Only the related single substitution analogues gave some cross-stimulation of T cells primed with the double substitutions. Thus, for a total of 18 different analogues, the T-cell repertoire of the B10 mouse was able to provide a unique response to every one. We conclude from these experiments that the functional T-cell repertoire is quite extensive.

This degree of diversity in the T-cell antigen receptor is somewhat surprising given the constraints imposed on the repertoire at both the genetic and cellular levels. First, although the *V*, *D* and *J* genetic elements of the T-cell receptor are as numerous as in B cells and joined with similar enzymatic machinery, thus giving rise to comparable amounts of combinatorial and junctional diversity, T cells do not appear to manifest a somatic hypermutation mechanism, which can increase diversity in B cells after antigen recognition has occurred⁶. Second, the T-cell population undergoes a positive selection step during development in the thymus which skews the repertoire toward recognition of self MHC-encoded molecules⁷. During this process many potential receptor combinations are not selected or expanded for use in the peripheral T-cell population. Thus, one might have expected the mature T-cell repertoire to be far more limited than the B-cell repertoire. Although this may still be the case, the failure so far to find many holes in the T-cell repertoire would suggest that it is diverse enough to respond uniquely to most antigenic challenges.

What then accounts for *Ir* gene nonresponsiveness? Any given MHC-encoded molecule has a high probability of not working well with a given antigenic determinant. For example, the MHC class II molecule $A_{\beta}^b:A_{\alpha}^b$ of the B10 mouse strain used in this study results in low or nonresponsiveness with about 65% of all tested immunogens under *Ir* gene control¹. It is not clear how much of this nonresponsiveness is due to holes in the repertoire and how much to failed physical interaction with $A_{\beta}^b:A_{\alpha}^b$. However, in the present study this question does not arise, as the peptides used are very similar in the regions that interact with the MHC-encoded molecule. Every analogue we made elicited a unique population of T-cell clones having greatest specificity for the immunizing peptide. These findings greatly extend the early studies of Levine *et al.*⁸ and the more recent studies of Mills *et al.*⁹, Gammon *et al.*¹⁰ and ourselves¹¹ which examined only 2 or 3 analogues of several antigenic peptides. The current experiments failed to find a single hole in the T-cell repertoire for one antigenic determinant despite 18 different probes with changes located at two different sites on the peptide. If this observation is substantiated for other antigenic determinants, it would suggest that the major *Ir* gene constraint on observed T-cell immune responsiveness lies at the level of the MHC-encoded molecule–antigen interaction, rather than in the T-cell repertoire².

We thank Ms Cindy French for HPLC purification of the peptides, Drs Ronald Germain, William Paul, Jonathan Ashwell, Barbara Fox, Marc Jenkins and Kathleen McCoy for helpful suggestions and critical review of the manuscript, and Ms Shirley Starnes for editorial assistance.

Received 16 September; accepted 2 December 1986.

1. Schwartz, R. H. *Adv. Immun.* **38**, 31–201 (1986).
2. Schwartz, R. H. *Nature* **317**, 284–285 (1985).
3. Schwartz, R. H. *A. Rev. Immun.* **3**, 237–261 (1985).
4. Schwartz, R. H. *et al. J. Immun.* **135**, 2598–2608 (1985).
5. Suzuki, G. & Schwartz, R. H. *J. Immun.* **136**, 230–239 (1986).
6. Fink, P. J., Matis, L. A., McElligott, D. L., Bookman, M. & Hedrick, S. M. *Nature* **321**, 219–226 (1986).
7. Schwartz, R. H. in *Fundamental Immunology* (ed. Paul, W. E.) 379–438 (Raven, New York, 1984).
8. Levine, B. B., Ojeda, A. & Benacerraf, B. *Nature* **200**, 544–546 (1963).
9. Mills, K. H. G., Skehel, J. J. & Thomas, D. B. *J. exp. Med.* **163**, 1477–1490 (1986).
10. Gammon, G. *et al. Nature* **319**, 413–415 (1986).
11. Hansburg, D., Fairwell, T., Schwartz, R. H. & Appella, E. *J. Immun.* **131**, 319–324 (1983).
12. Schwartz, R. H., Horton, C. L. & Paul, W. E. *J. exp. Med.* **145**, 327–343 (1977).

Unusual intron in the immunoglobulin domain of the newly isolated murine CD4 (L3T4) gene

Dan R. Littman & Sonya N. Gettner

Department of Microbiology and Immunology,
University of California, San Francisco, California 94143, USA

The T-cell surface glycoprotein, CD4, is expressed predominantly on helper T cells and is thought to play a major role in cell-cell interactions¹⁻⁵. Monoclonal antibodies against CD4 have been shown to block numerous T-cell functions^{1,3,4}; moreover, recent results suggest that the CD4 molecule may be involved in transmembrane signal transduction^{6,7}. The human CD4 glycoprotein has also been shown to form at least part of the receptor for the AIDS virus, HIV-1 (refs 8-10). Elucidation of the functions of CD4 will be facilitated by the ability to manipulate the protein by genetic means. Because the mouse system is well suited for a variety of functional studies, we have isolated, sequenced and expressed cDNA clones encoding the murine CD4 (L3T4) glycoprotein. Comparison of the mouse and human CD4 sequences reveals striking evolutionary conservation of the cytoplasmic domain, suggesting that this region is essential for CD4 function. In addition, both the human and mouse CD4 gene contain a large intron in the coding region of the V-like domain. As no other members of the immunoglobulin gene superfamily have been shown to contain similarly placed introns, this finding may have important implications regarding the evolution of this gene family in particular and of introns in general.

Murine CD4 complementary DNAs were isolated from a thymocyte library prepared in λ gt10 and from a clonal T-cell library prepared in pCD (ref. 11). The libraries were screened under conditions of low stringency with the 3-kilobase (kb) human CD4 cDNA¹². Two positive clones, the 2.1-kb p3C in λ gt10 and the 3.1-kb L3T4.25 in pCD, were subjected to further analysis (Figs 1a and 2a). To confirm that we had isolated cDNAs encoding murine CD4, as defined by the monoclonal antibody GK1.5, we co-transformed *Ltk*⁻ cells with the pCD.L3T4.25 plasmid and pTK. *Tk*⁺ colonies were shown to express surface L3T4 antigen both by erythrocyte rosetting¹³ and by FACS (fluorescence-activated cell sorter) analysis (Fig. 1b), indicating that the region of murine CD4 recognized by GK1.5 is encoded by the L3T4.25 insert.

In the sequence of the full-length murine CD4 cDNA (Fig. 2), an open reading frame begins at nucleotide 170 and encodes a protein of 457 amino-acid residues. Its predicted relative molecular mass, not including a putative signal peptide, is 48,335. The leader sequence is followed by a domain of 110 residues which, as in T4, shows extensive similarity with variable regions of immunoglobulin light chains (Fig. 3). There is also significant similarity with the α chain of T-cell receptor (Fig. 3). Overall homology between murine CD4 and representative V_{κ} and V_{α} domains is 26% and 23%, respectively. Cysteines at positions +16 and +86 of murine CD4 correspond to cysteines involved in intrachain disulphide bond formation in the immunoglobulins. In our previous analysis, weak homology was found between a region beyond the V-like domain of human CD4 and immunoglobulin and T-cell receptor J segments¹². There is also weak homology between residues 102-108 of murine CD4 and J segments, with the consensus J region motif F-G-X-G-T partially conserved as F-S-P-G-T in murine CD4. However, this J-like sequence is shifted relative to the position of the J-like sequence in human CD4; moreover, the human J-like segment is interrupted by an intron. It is therefore unlikely that these sequences have resulted from evolutionary conservation of J segments.

The Ig-like domain of murine CD4 is followed by 270 residues

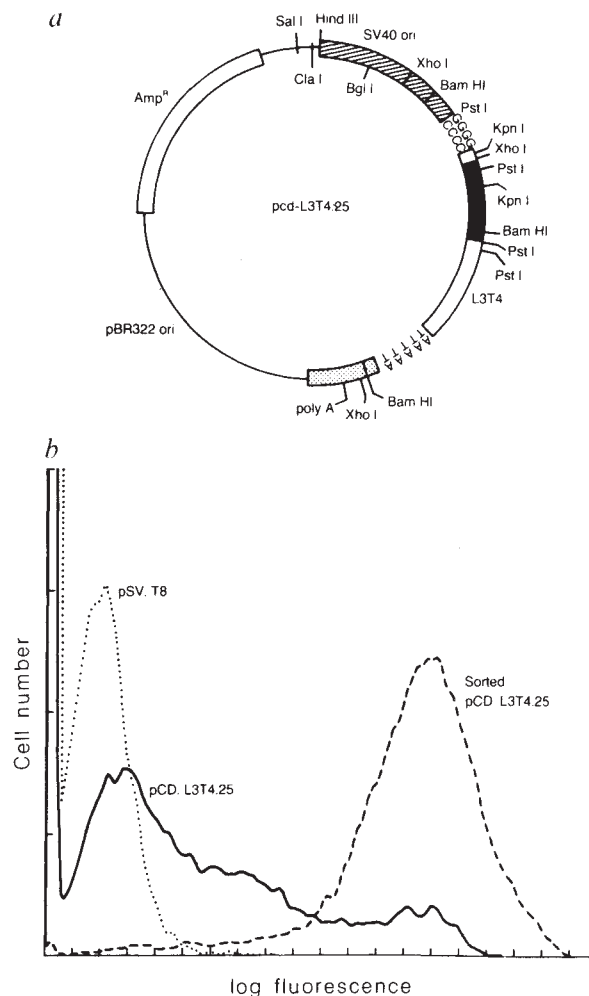
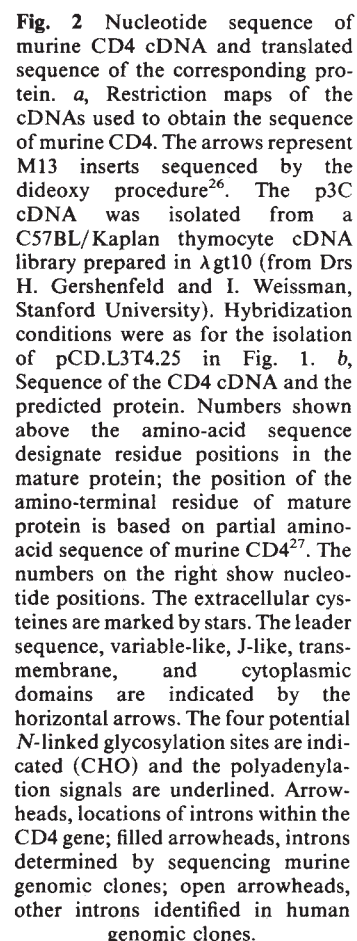


Fig. 1 Expression of murine CD4 on the surface of L cells transfected with pTK and pCD.L3T4.25. *a*, Restriction map of pCD.L3T4.25. The plasmid was isolated from a cDNA library prepared from the T-cell clone 5C (ref. 11) and obtained from Dr K. I. Arai (DNAX). The library was screened with the full-length human T4 cDNA probe¹². Hybridization was at 42 °C in 30% formamide, 5 × SSCPE, 10% dextran sulphate, 1 × Denhardt's, and 100 μ g ml⁻¹ denatured salmon sperm DNA; the filters were washed in 2 × SSC, 0.1% SDS at 50 °C. *b*, Cytofluorometric analysis of transfected L cells stained with the monoclonal antibody GK1.5. *La*⁻ cells were co-transformed as described previously using 100 ng pTK and 2 μ g pCD.L3T4.25 (solid line) or 2 μ g control CD8 plasmid (dotted line). HAT-resistant colonies were pooled and were stained with phycoerythrin-conjugated GK1.5 (Becton-Dickinson Monoclonals). The cells were analysed on a FACS IV cell sorter. Stable transformants expressing high levels of murine CD4 were enriched through one round of cell sorting, and are included in the figure.

having no similarity to other sequences in the data bank. This sequence includes four other extracellular cysteines which are conserved between mouse and human CD4. Two of these, Cys 133 and Cys 162, are present within an 80-residue domain encoded by a single exon which also contains the J-like residues (Fig. 2b); and the others are at positions 302 and 344, in the membrane-proximal region of CD4 which is also encoded by a separate exon in the human CD4 gene. It has been shown that the three pairs of cysteine residues form disulphide bonds within the individual domains¹⁴. The extracellular domains of murine CD4 contain four potential sites for N-linked glycosylation, Asn-X-Thr where X represents any amino acid. One of these, Asn 161, may not be used if Cys 162 is involved in disulphide bonding; and Asn 376, immediately proximal to the transmembrane domain, may also not be accessible to glycosylation. The



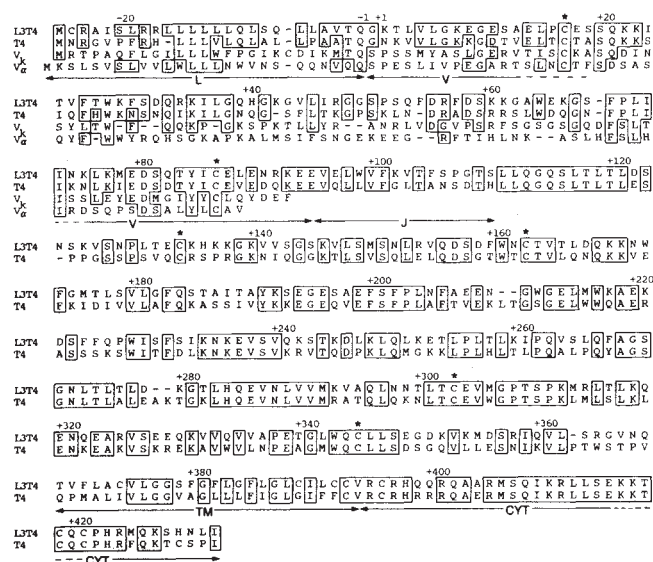


Fig. 3 Alignment of murine CD4 (L3T4) protein sequence with the sequences of human CD4 (T4) (ref. 12) and of representative murine kappa²⁸ and alpha²⁹ chain variable regions. Best matches were obtained by search of the Dayhoff protein sequence library.

remaining sites, Asn 272 and Asn 297, correspond to the two glycosylation sites in human CD4. The extracellular domains of murine CD4 are followed by a putative transmembrane region, consisting of 25 largely hydrophobic residues, and by a highly charged cytoplasmic domain of 38 amino-acid residues. Whereas the overall homology between the human and murine CD4 molecules is 55%, homology between the cytoplasmic domains is 80%, with a stretch of identity in 29 out of 32 residues (Fig. 3).

In the sequence of CD4 genomic clones, we found an unexpected intron separating the two halves of the V-like domain of both human and murine CD4 (Fig. 4). In other members of the immunoglobulin gene family, the Ig domains are encoded by single exons of ~300 nucleotides¹⁵. The high degree of homology between immunoglobulin and T-cell receptor variable regions and CD4 suggests that these genes resulted from a duplication event relatively late in the evolution of the Ig family. Our observation is therefore best explained by the insertion of an intron in a V-domain exon during the evolution of CD4, before human-rodent divergence. This interpretation differs from the more widely accepted hypothesis^{16,17}, supported by recent studies on the triosephosphate isomerase gene^{18,19}, that introns are evolutionary remnants formed by the shuffling of exons. According to the exon shuffling interpretation, the CD4 gene would correspond to an ancestral Ig domain from which the other members of the Ig superfamily evolved following duplication and loss of the intron. The finding that the position of the intron in CD4 corresponds to a hinge region separating the C' and D β strands of the Ig fold²⁰ supports this view; the non-random position of this intron is consistent with the finding that exons encoding discrete protein domains are often separated by intron sequences²¹, as would be expected from exon migration. However, this interpretation is unlikely to be correct in light of the considerable similarity between CD4 and the other members of the Ig superfamily. Strong evolutionary pressure would be required to conserve such extensive sequence homology. A third possibility is that the Ig fold of CD4 resulted from convergent evolution; this is also very unlikely, because much lower homology can be tolerated in the formation of an Ig fold²². A study of the structure of the CD4 gene in lower species may provide some hint as to which of these hypotheses is most likely to explain this observation.

In the periphery, the CD4 and CD8 glycoproteins are distributed on functionally distinct subpopulations of T cells^{2,4}.

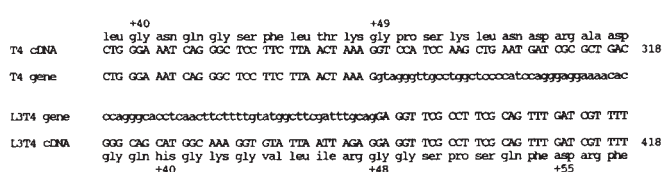


Fig. 4 Nucleotide sequence of the V-like region intron-exon boundaries within the murine (L3T4) and human (T4) CD4 genes. The numbers designate the amino-acid residue position in the mature protein. Genomic libraries of human fetal liver³⁰ and BALB/c mammary tumour (gift of T. Fong) were screened with full-length human or murine CD4 cDNA probes. Inserts were subcloned into M13 and sequenced by the dideoxy procedure²⁶.

CD4⁺ cells have T-cell antigen receptors restricted to recognize antigen in association with self class II MHC molecules^{3,4}; CD8⁺ cells recognize antigen and class I MHC molecules²³. It has therefore been suggested that CD4 and CD8 bind directly to non-polymorphic domains on the cognate major histocompatibility complex (MHC) molecules to increase the avidity of cell-cell interactions^{5,23}. Although CD4 may act simply as an adhesion molecule^{5,24,25}, recent data suggest that it may also be involved in transmembrane signalling^{6,7}. In these experiments, T-cell activation with anti-CD4 antibodies or with concanavalin A is abrogated by treatment with anti-CD4 antibodies in the absence of any accessory cells. The striking homology between the human and mouse CD4 cytoplasmic domains further supports the notion that the CD4 glycoprotein is involved in mediating a signal across the plasma membrane. Although the CD4 molecule has not yet been shown to be associated with other T-cell membrane proteins, it is quite likely that it interacts with some component of the T3/Ti complex in the course of T-cell activation. As both the T-cell receptor and the CD4 molecule are thought to bind components of class II MHC molecules on target cells, this may be a mechanism for facilitating the bridging of these T-cell membrane glycoproteins during cell-cell interactions, resulting in a transmembrane signal. Future studies in reconstituted systems will be required to elucidate the mechanism of action of CD4.

We thank Paul Dazin for the cytofluorometric analysis, Irv Weissman and Ken-Ichi Arai for providing cDNA libraries, Elliot Epner for assistance with the genomic subcloning, and Nils Lonberg for helpful discussions. This work was supported by NIH grant AI23513 and by a Searle Scholar Award to D.L.

Received 10 September; accepted 26 November 1986.

- Webb, M., Mason, D. W. & Williams, A. F. *Nature* **282**, 841-843 (1979).
- Reinherz, E. L. & Schlossman, S. F. *Cell* **19**, 821-827 (1980).
- Krensky, A. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2365-2369 (1982).
- Swain, S. L., Dialynas, D., Fitch, F. W. & English, M. J. *Immun.* **132**, 1118-1123 (1984).
- Marrack, P. et al. *J. exp. Med.* **158**, 1077-1091 (1983).
- Wassmer, P., Chan, C., Logdberg, L. & Shevach, E. M. *J. Immun.* **135**, 2237-2242 (1985).
- Bank, I. & Chess, L. *J. exp. Med.* **162**, 1294-1303 (1985).
- Dagleish, A. et al. *Nature* **312**, 763-767 (1984).
- Klatzmann, D. et al. *Nature* **312**, 767-768 (1984).
- McDougal, J. S. et al. *Science* **231**, 382-385 (1986).
- Hiraki, D. D. et al. *J. Immun.* **136**, 4291-4296 (1986).
- Maddon, P. J. et al. *Cell* **42**, 93-104 (1985).
- Littman, D. R. et al. *Cell* **40**, 237-246 (1985).
- Classon, B. J., Tsagaratos, J., McKenzie, I. F. C. & Walker, I. D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4499-4503 (1986).
- Hood, L., Kronenberg, M. & Hunkapiller, T. *Cell* **40**, 225-229 (1985).
- Doolittle, R. *Nature* **272**, 581-582 (1978).
- Gilbert, W. *Nature* **271**, 501 (1978).
- Marchionni, M. & Gilbert, W. *Cell* **46**, 133-141 (1986).
- McKnight, G. L., O'Hara, P. J. & Parker, M. L. *Cell* **46**, 143-147 (1986).
- Amzel, L. M. & Poljak, R. J. *A. Rev. Immun.* **48**, 961-997 (1979).
- Go, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1964-1968 (1983).
- Williams, A. F. *Immun. Today* **5**, 219-221.
- Swain, S. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7101-7105 (1981).
- Greenstein, J. L., Kappler, J., Marrack, P., & Burakoff, S. J. *J. exp. Med.* **159**, 1213-1224 (1984).
- Biddison, W. E., Rao, P. E., Talle, M. A., Goldstein, G. & Shaw, S. J. *J. exp. Med.* **159**, 783-797 (1984).
- Sanger, F., Nicklen, S. & Coulson, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Classon, B. J. et al. *Immunogenetics* **23**, 129-132.
- Altenburger, W., Steinmetz, M. & Zachau, H. G. *Nature* **287**, 603-607 (1980).
- Chien, Y. et al. *Nature* **312**, 31-35 (1984).
- Maniatis, T. et al. *Cell* **15**, 687-701 (1978).

Activation of sodium-proton exchange is a prerequisite for Ca^{2+} mobilization in human platelets

W. Siffert* & J. W. N. Akkerman†

* Institut für Physiologie, Ruhr-Universität, D-4630 Bochum (FRG)

† Department of Haematology, University Hospital, NL-3511 GV Utrecht, The Netherlands

Stimulated platelets take up sodium ions^{1,2} and release hydrogen ions³ due to activation of Na^+/H^+ exchange^{4,5} resulting in cytoplasmic alkalization⁶. Suppression of Na^+/H^+ exchange either by removal of extracellular Na^+ or by application of amiloride inhibits shape change, secretion of granule contents and aggregation^{5,7-9}. The data we present here indicate that inhibition of this transport by ethylisopropyl-amiloride or by lowering extracellular sodium reduces or even completely suppresses the rise in cytoplasmic free Ca^{2+} concentration that is essential for platelet aggregation in response to thrombin. We also demonstrate that cytoplasmic alkalization produced by exposure to the ionophore monensin sensitizes the human platelet response to stimulation by thrombin resulting in enhanced Ca^{2+} mobilization and aggregability. We conclude that an increase in intracellular pH evoked by activation of Na^+/H^+ counter transport is an important signal in stimulus-response coupling and forms an essential step in the cascade of events required to increase cytoplasmic free Ca^{2+} in platelets.

It has been demonstrated for a variety of cells and tissues that receptor occupancy leads to the rapid hydrolysis of phosphatidylinositol 4,5-bisphosphate ($\text{PtdIns } 4,5 \text{ P}_2$) with subsequent formation of 1,2-diacylglycerol (DG) and inositol 1,4,5-trisphosphate (InsP_3). Whereas InsP_3 appears to be an intracellular messenger for Ca^{2+} release from internal stores, DG has been shown to induce Na^+/H^+ exchange via activation of protein kinase C^{10-12} and the resultant elevation of intracellular pH (pH_i) appears to be intimately linked with stimulus-response coupling^{13,14}. The mechanisms by which an increase in pH_i can promote cellular functions have however, remained obscure.

The aim of the present study was to obtain further insights into the role of Na^+/H^+ exchange for cell activation using the human platelet as a model. Our investigation focused on the impact of Na^+/H^+ countertransport on basic platelet functions, that is, mobilization of Ca^{2+} and platelet aggregation.

In a first set of experiments the effects of replacing extracellular Na^+ (Na_o^+) on thrombin-induced Ca^{2+} mobilization was investigated (Fig. 1) using the fluorescent dye quin-2. In medium containing 1 mM Ca^{2+} and 140 mM Na_o^+ (Fig. 1a) the addition of thrombin significantly raised the concentration of cytoplasmic free Ca^{2+} , $[\text{Ca}^{2+}]_i$, on average from 102 ± 27 to 917 ± 210 nM ($n=9$; $\pm$ s.d.), whereas at 0.5 mM Na_o^+ (Fig. 1b) only a small increase in $[\text{Ca}^{2+}]_i$ was observed, from an average of 94 ± 25 to 183 ± 20 nM ($n=7$; $\pm$ s.d.). In addition, the time required for the quin-2 signal to reach its peak value was distinctly prolonged. In Ca^{2+} -free medium (2 mM EGTA) containing 140 mM Na_o^+ (Fig. 1c) $[\text{Ca}^{2+}]_i$ rose from 106 ± 31 to 247 ± 92 nM ($n=7$; $\pm$ s.d.) upon addition of thrombin, but at low Na_o^+ (Fig. 1d) $[\text{Ca}^{2+}]_i$ remained at its resting level of 108 ± 14 nM ($n=6$; $\pm$ s.d.). Identical results were obtained when NaCl was replaced by other substitutes such as bis(2-hydroxyethyl)dimethylammonium or *N*-methyl-D-glucamine.

It thus appears that suppression of Na^+/H^+ exchange by removal of Na_o^+ can completely inhibit increases in $[\text{Ca}^{2+}]_i$ in conditions in which an influx of Ca^{2+} from the suspending medium is negligibly small and, thus, changes in the fluorescence of quin-2 reflect mobilization of Ca^{2+} from internal stores. A substantial inhibition by low Na_o^+ is also observed in the presence of 1 mM Ca_o^{2+} , suggesting that Na^+/H^+ exchange may take

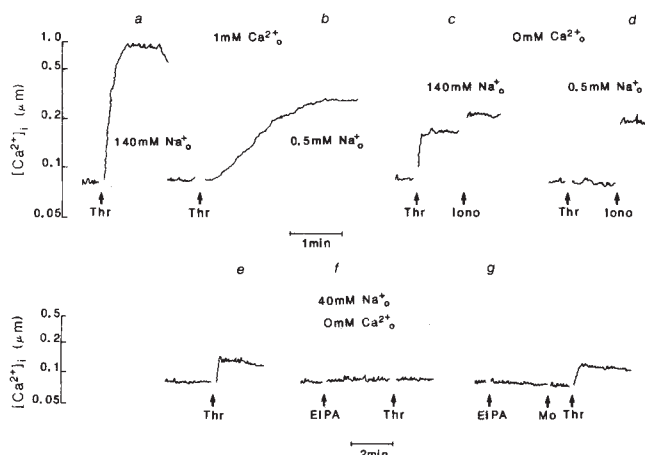


Fig. 1 Effect of external Na^+ or ethylisopropyl-amiloride (EIPA) on thrombin-induced Ca^{2+} mobilization in human platelets. The traces reflect the increase in $[\text{Ca}^{2+}]_i$ as evaluated from the changes in fluorescence of quin-2-loaded platelets. Each trace is representative for at least six similar experiments performed in three different preparations. a, b, The effect of 0.1 U ml^{-1} thrombin (Sigma) on $[\text{Ca}^{2+}]_i$ in platelets suspended in normal (140 mM, a) or low (0.5 mM, b) Na^+ -medium in the presence of 1 mM Ca_o^{2+} . c, d, Response of platelets at 140 mM (c) and 0.5 mM $[\text{Na}^+]_o$ (d). 2 mM EGTA was present to set $[\text{Ca}^{2+}]_o$ below 100 nM. After stimulation by thrombin (Thr) (0.1 U ml^{-1}) 100 nM ionomycin (Iono) was added to indicate the total amount of Ca^{2+} that could be released from internal stores. It is obvious that lowering $[\text{Na}^+]_o$ to 0.5 mM results in a distinct inhibition of thrombin-induced Ca^{2+} mobilization both at 1 mM and 0 mM $[\text{Ca}^{2+}]_o$ (b, d) as compared with controls (a, c). Effect of EIPA on Ca^{2+} mobilization. e, Control (0.1 U ml^{-1} thrombin); f, addition of 40 μM EIPA before thrombin; g, preincubation with 40 μM EIPA followed by monensin (Mo) (10 μM). Experiments in e–g were performed in Ca^{2+} -free medium (2 mM EGTA) at 40 mM that leads to a smaller quin-2 signal than at 140 mM Na_o^+ (see also Fig. 2). This concentration was chosen to allow complete inhibition of Na^+/H^+ exchange by 40 μM EIPA, because this compound competes with Na_o^+ (ref. 16). Higher concentrations of EIPA ($>50 \mu\text{M}$) disturbed the quin-2 signals.

Methods. Freshly drawn citrated blood was centrifuged at 200g for 15 min to obtain platelet-rich plasma. The platelet-rich plasma was then incubated for 20 min with 20 μM quin-acetoxymethyl ester (Sigma) at room temperature. After addition of $0.1 \mu\text{M}$ forskolin the quin-2-loaded cells were centrifuged at 700g for 20 min. The pellet was resuspended into 1 ml of HEPES buffer consisting of 140 mM choline chloride (Merck), 5 mM KCl, 1 mM MgSO_4 , 10 mM HEPES, 5 mM glucose pH 7.5 and $0.1 \mu\text{M}$ forskolin and 10 μM EGTA were added. Finally the suspension was passed through a Sepharose 2B column pre-equilibrated with the same buffer (presence of 10 μM EGTA, no forskolin) to remove extraneous dye and extracellular Na^+ (Na_o^+). The cells were then incubated at 37°C in media containing various $[\text{Na}^+]_o$ which was achieved by mixing various volumes of HEPES buffer containing either 140 mM Na^+ or 140 mM choline $^+$. CaCl_2 1 mM or 2 mM EGTA was added as required, and the cells (1×10^8 per ml) were pre-equilibrated in a cuvette at 37°C for 5 min before addition of thrombin. The quin-2 fluorescence was measured in a Perkin Elmer MPF-3 spectrofluorometer, with excitation at 340 nm and emission at 490 nm essentially as described by Tsien *et al.*²⁶. EIPA was a gift from Drs Th. Friedrich and G. Burckhardt. A 10 mM stock was prepared in 50% (v/v) DMSO.

part in the regulation of the influx of Ca^{2+} . It is unlikely that these observations are due to disturbances of Ca^{2+} homeostasis by removal of Na_o^+ in unstimulated platelets, because $[\text{Ca}^{2+}]_i$ in resting platelets did not depend on the presence of Na^+ in the suspension medium and the addition of the Ca^{2+} ionophore ionomycin to platelets suspended in Ca^{2+} -free medium produced identical Ca^{2+} signals at normal or low Na_o^+ (Fig. 1c, d), which indicates that the total amount of Ca^{2+} releasable from internal stores was unaffected. In addition, it has recently been

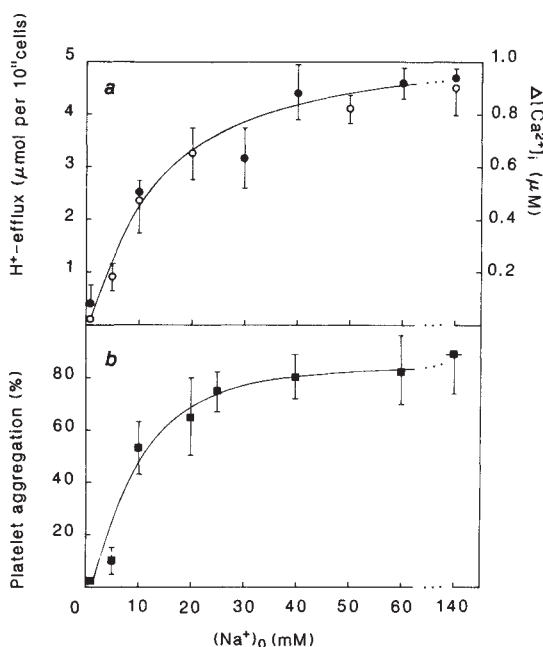


Fig. 2 Dependence on external Na^+ of proton efflux, Ca^{2+} mobilization and aggregation of thrombin-stimulated platelets. *a*, The Na^+ -dependence of Na^+/H^+ exchange in thrombin-activated platelets was measured by following the amiloride-sensitive (1 mM amiloride) H^+ extrusion in a poorly buffered medium as previously described⁴. Briefly, platelets were pelleted from platelet-rich plasma and resuspended in a medium consisting of 140 mM NaCl, 2.7 mM KCl, 0.18 mM KH_2PO_4 , 5 mM glucose pH 7.4. From the observed changes in medium pH and the buffer capacity the total amount of hydrogen ions released could be calculated and is expressed as μ mol H^+ per 10^{11} cells (left scale). Measurements were performed at various $[Na^+]_0$ after isotonic replacement of Na^+ by choline⁺. The cell count was 2×10^8 per ml and the stimulus thrombin at 0.5 U ml^{-1} . The points (open circles, $\circ$) represent means of at least three determinations ($\pm$ s.d.). Ca^{2+} mobilization in response to thrombin (0.1 U ml^{-1}) as a function of $[Na^+]_0$ was determined in quin-2-loaded platelets at 1 mM external Ca^{2+} ($\bullet$, right scale). Experimental details are identical to those described in the legend to Fig. 1. Means of at least four determinations from three different preparations ($\pm$ s.d.). *b*, Requirement for external Na^+ of platelet aggregation. Pelleted platelets were resuspended in HEPES buffer containing additionally 1 mM Ca^{2+} and albumine. Aggregation was measured by a standard turbidimetric technique²⁷. $[Na^+]_0$ was lowered by isotonic replacement of Na^+ with choline⁺. Means of at least three determinations from two different preparations ($\pm$ s.d.). H^+ release, Ca^{2+} mobilization and aggregation show an almost identical dependence on Na^+ , strongly supporting the idea that activation of Na^+/H^+ exchange is an essential part of platelet function.

demonstrated that the total platelet Ca^{2+} content is independent from the transmembrane Na^+ gradient¹⁵. The amiloride derivative ethylisopropyl-amiloride (EIPA), a potent inhibitor of Na^+/H^+ exchange¹⁶, also suppressed Ca^{2+} mobilization at 0 mM Ca^{2+}_0 (Fig. 1e, f). The inhibitory action of this compound was bypassed after artificial alkalization of the platelet cytoplasm by use of the Na^+/H^+ ionophore monensin (Fig. 1g). A substantial inhibition of Ca^{2+} mobilization was also observed at 1 mM Ca^{2+}_0 (data not shown). These findings add strong support to the idea that Na^+/H^+ exchange is required for Ca^{2+} mobilization.

In a similar experimental approach we determined the requirement for Na^+ of amiloride-sensitive H^+ efflux, Ca^{2+} mobilization and aggregation in thrombin-stimulated platelets (Fig. 2). Thrombin-stimulated H^+ efflux depended on Na^+ (Fig. 2a) confirming earlier reports that the transmembrane Na^+ gradient

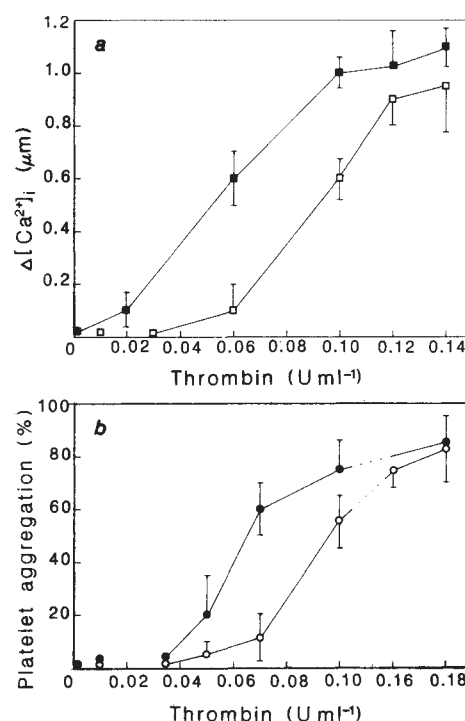


Fig. 3 Effect of pretreatment with monensin on Ca^{2+} mobilization and platelet aggregation. *a*, Changes in $[Ca^{2+}]_i$ (μ M) measured at various concentrations of thrombin (U ml^{-1}). Experiments were performed in quin-2-loaded platelets (for details see Fig. 1) suspended in 140 mM Na^+ -medium ($[Ca^{2+}]_0 = 1 \text{ mM}$). $\square$, Controls; $\blacksquare$, platelets, pretreated with 10 μ M monensin (Sigma) for 1 min. Means of at least three determinations from two different preparations ($\pm$ s.d.). *b*, Thrombin-induced platelet aggregation before ($\circ$) and after treatment with 10 μ M monensin for 1 min ($\bullet$). Platelets were suspended in 140 mM Na^+ -medium. Means of at least four determinations from three different preparations ($\pm$ s.d.). Exposure of platelets to monensin, that is, increasing intracellular pH sensitizes platelets towards stimulation by thrombin as can be seen from increased Ca^{2+} mobilization and increased aggregability.

is the driving force for H^+ release⁴. The dependence on Na^+ of Ca^{2+} mobilization (Fig. 2a) and platelet aggregation (Fig. 2b) was almost identical with that of H^+ release, which adds support to our hypothesis that the inhibitory effect of Na^+ removal on both functions specifically derives from inhibition of Na^+/H^+ exchange. Half maximal activation occurred at about 10 mM Na^+ , which is in good agreement with values reported for most of the Na^+/H^+ carriers hitherto described¹⁷.

The physiological significance of cytoplasmic alkalization could also be demonstrated in experiments in which Na^+/H^+ exchange was artificially induced by the ionophore monensin. Pre-treatment of platelets with this compound resulted in a distinct enhancement of Ca^{2+} mobilization (Fig. 3a) and platelet aggregation (Fig. 3b) in response to thrombin. It is evident, however, that an increase in pH_i alone does not evoke these responses because application of monensin alone did not raise $[Ca^{2+}]_i$ nor could it induce platelet aggregation in the absence of thrombin.

To reconfirm that thrombin actually increases pH_i in human platelets via amiloride-sensitive Na^+/H^+ exchange as previously suggested⁶, we have measured pH_i in thrombin-stimulated platelets after loading the cells with the fluorescent, pH-sensitive dye 6-carboxyfluorescein-diacetate (6-CF). Upon addition of thrombin (0.1 U ml^{-1}), pH_i increased by 0.17 ± 0.05 units ($n = 7 \pm$ s.d.; Fig. 4a), whereas no increase in pH_i was observed after reducing Na^+ to 5 mM (Fig. 4b) or after blocking Na^+/H^+ exchange by 100 μ M EIPA (Fig. 4c).

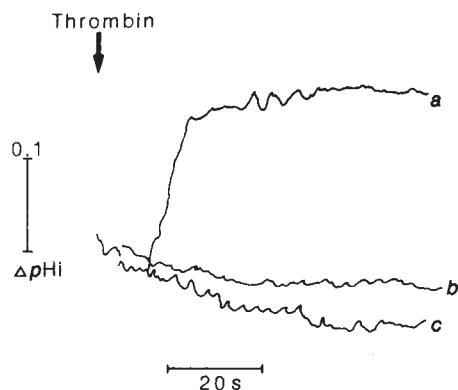


Fig. 4 Effect of thrombin on cytoplasmic pH_i in human platelets. The traces reflect the changes in fluorescence observed in platelets loaded with the pH -sensitive dye 6-carboxyfluorescein diacetate (6-CF). *a*, Response to thrombin (0.1 U ml^{-1}) in 140 mM Na^+ medium; *b*, Na^+ reduced to 5 mM ; *c*, 140 mM Na^+ plus $100 \mu\text{M}$ EIPA. The dependence on Na^+ and inhibition by EIPA of the thrombin-induced increase in fluorescence argues in favour of a Na^+/H^+ exchange-mediated increase in pH_i . Each trace is representative for at least 6 experiments from 4 different preparations. **Methods.** Platelets were labelled with 6-CF using a slightly modified preparation as described by Horne *et al.*⁶. Platelets were pelleted from platelet-rich plasma after addition of $1 \mu\text{M}$ forskolin and $10 \mu\text{M}$ EGTA by centrifugation at $600g$ for 15 min . The pellet was resuspended in HEPES buffer (140 mM NaCl , 5 mM KCl , 0.1% gelatine (w/v), 1 mM MgSO_4 , 10 mM HEPES , $15 \text{ mM citric acid}$, 5 mM glucose $pH 7.5$) and incubated with $20 \mu\text{M}$ 6-CF (added from a 20 mM stock in DMSO) for 15 min at 37°C . Thereafter platelets were pelleted again and washed once in HEPES buffer followed by resuspension in the same medium (no forskolin). The labelled platelets were kept at 37°C at a concentration of 5×10^9 cells per ml and these cells were fully responsive to thrombin (0.03 U ml^{-1}) or ADP ($5 \mu\text{M}$) as tested by aggregometry. Immediately before starting the experiments platelets were diluted to 5×10^7 cells per ml in HEPES buffer (see above) containing either 140 mM Na^+ or 140 mM choline . λ_{exc} was 492 nm , λ_{em} was 518 nm . Leakage of dye was negligibly small during the experiments, so that no correction had to be made. Absolute pH_i values were obtained by the nigericin/KCl method⁶. In resting cells pH_i was found to be $7.1 (\pm 0.05; n = 8 \pm \text{s.d.})$. After addition of thrombin the fluorescence signal dropped unspecifically (which was ascertained by addition of equal amounts of thrombin-free buffer) due to altered light scattering through the cell suspensions. For clearness of presentation, the traces shown in Fig. 4 therefore reflect the fluorescence after addition of agonists. 6-CF was a gift of Dr I. Schulz (Frankfurt).

Stimulation of platelets by thrombin results in a rapid phosphodiesteratic cleavage of $\text{PtdIns } 4,5 \text{ P}_2$ (ref. 18). One of the products of this reaction, DG, combines with the protein kinase C which in turn stimulates the phosphorylation of a protein of $47,000$ (47K) molecular mass¹⁹ that may be directly related to the release reaction. In addition, activation of the protein kinase C initiates Na^+/H^+ exchange in many cells¹¹ and this may also hold true for human platelets, as we find that Na^+/H^+ exchange can be activated by a phorbol ester ($12\text{-O-tetradecanoyl phorbol-13-acetate}$, TPA) and exogenous diacylglycerol ($1\text{-oleoyl-2-acetyl-glycerol}$; OAG; W.S. unpublished data). The other product, InsP_3 , is a putative signal for the release of Ca^{2+} from non-mitochondrial intracellular stores^{11,20} and has also been shown to release Ca^{2+} from intracellular membranes derived from human platelets²¹⁻²³. Our results, however, suggest, that both signals, an increase in intracellular InsP_3 concentration and an increase in pH_i are simultaneously required for Ca^{2+} mobilization. Increases in pH_i alone, on the other hand do not evoke Ca^{2+} mobilization, because monensin (this study) as well as TPA and OAG²⁴ do not produce changes in $[\text{Ca}^{2+}]_i$, at least as estimated from determinations in quin-2-loaded platelets.

Recently Brass and Joseph²⁵ demonstrated that increasing the extracellular pH from $pH 7.1$ to 7.4 lowered the apparent K_M of InsP_3 towards the Ca^{2+} stores by about 40% in permeabilized platelets providing additional evidence for a role of cytoplasmic alkalinization in Ca^{2+} mobilization.

We thank Mrs T. Riemens and Mrs G. Siffert for technical assistance, Drs T. Friedrich and G. Burckhardt for the generous gift of EIPA, and Dr I. Schulz for the supply of 6-CF. We are also indebted to Drs A. Wegner (Bochum) and W. Moolenaar (Utrecht) for helpful discussions. This work was supported in part by the Deutsche Forschungsgemeinschaft (grant Sche 46-5/1).

Received 30 July; accepted 31 October 1986.

1. Feinberg, H. *et al. Biochim. biophys. Acta* **470**, 317-324 (1977).
2. Sandler, W. C., Le Breton, G. C. & Feinberg, H. *Biochim. biophys. Acta* **600**, 448-455 (1980).
3. Akkerman, J. W. N. & Holmsen, H. *Blood* **57**, 956-966 (1981).
4. Siffert, W., Fox, G., Mückenhoff, K. & Scheid, P. *FEBS Lett.* **172**, 272-274 (1984).
5. Siffert, W., Mückenhoff, K. & Scheid, P. *Biochem. biophys. Res. Commun.* **125**, 1123-1128 (1984).
6. Horne, W. C., Norman, N. E., Schwartz, D. B. & Simons, E. R. *Eur. J. Biochem.* **120**, 295-302 (1981).
7. Horne, W. C. & Simons, E. R. *Thromb. Res.* **13**, 599-607 (1978).
8. Connolly, T. M. & Limbird, L. E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5320-5324 (1983).
9. Leven, R. M., Gonnella, P. A., Reeber, M. J. & Nachmias, V. T. *Thromb. Haemostas.* **49**, 230-234 (1983).
10. Berridge, M. J. *Biochem. J.* **220**, 345-360 (1984).
11. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315-321 (1984).
12. Macara, I. G. *Am. J. Physiol.* **248**, C3-C11 (1985).
13. Busa, W. B. & Nuccitelli, R. *Am. J. Physiol.* **246**, R409-R438 (1984).
14. Nuccitelli, R. & Deamer, D. W. *Intracellular pH: Its Measurement, Regulation and Utilization in Cellular Functions* (Liss, New York, 1982).
15. Brass, L. F. *J. Biol. Chem.* **259**, 12571-12575 (1984).
16. Vigne, P., Frelin, C., Cragoe, E. J. & Lazdunski, M. *Biochem. biophys. Res. Commun.* **116**, 86-90 (1983).
17. Aronson, P. S. *A. Rev. Physiol.* **47**, 545-560 (1985).
18. Agranoff, B. W., Murthy, P. & Seguin, P. *J. Biol. Chem.* **258**, 2076-2078 (1983).
19. Kawahara, Y., Takai, Y., Minakuchi, R., Sano, K. & Nishizuka, Y. *Biochem. biophys. Res. Commun.* **97**, 309-317 (1980).
20. Streb, H., Irvine, R. F., Berridge, M. J. & Schulz, I. *Nature* **306**, 67-68 (1983).
21. O'Rourke, F. A., Halenda, S. P., Zavoico, G. B. & Feinstein, M. B. *J. Biol. Chem.* **260**, 956-962 (1985).
22. Adunyah, S. E. & Dean, W. L. *Biochem. biophys. Res. Commun.* **128**, 1274-1280 (1985).
23. Authi, K. S. & Crawford, N. *Biochem. J.* **230**, 247-253 (1985).
24. Rink, T. J., Sanchez, A. & Hallam, T. J. *Nature* **305**, 317-319 (1983).
25. Brass, L. F. & Joseph, K. J. *J. Biol. Chem.* **260**, 15172-15179 (1985).
26. Tsien, R. Y., Pozzan, I. & Rink, T. J. *J. cell Biol.* **94**, 325-334 (1982).
27. Born, G. V. R. *Nature* **194**, 927-929 (1962).

Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease

Johannes Pohlner*, Roman Halter*, Konrad Beyreuther† & Thomas F. Meyer*

* Max-Planck-Institut für Biologie, Spemannstrasse 34, D-74 Tübingen, FRG

† Universität zu Köln, Institut für Genetik, D-5 Köln, FRG

Several human bacterial pathogens, including the Gram-negative diplococcus *Neisseria gonorrhoeae*, produce extracellular proteases that are specific for human immunoglobulin IgA₁ (refs 1, 2). Immunoglobulin A (IgA) proteases have been studied extensively and the genes of some species cloned in *Escherichia coli*³⁻⁷, but their role in pathogenesis remains unclear⁸. Recently we derived a DNA fragment of 5 kilobases (kb) from *N. gonorrhoeae* MS11 directing extracellular active enzyme in *E. coli*⁴. Although the mature enzyme of strain MS11 was shown to have a relative molecular mass of $106,000$ (M_r 106K) in gels⁴, the DNA sequence of this cloned fragment reveals a single gene coding for a 169K precursor of IgA protease. The precursor contains three functional domains, the amino-terminal leader which is assumed to initiate the inner membrane transport of the precursor, the protease, and

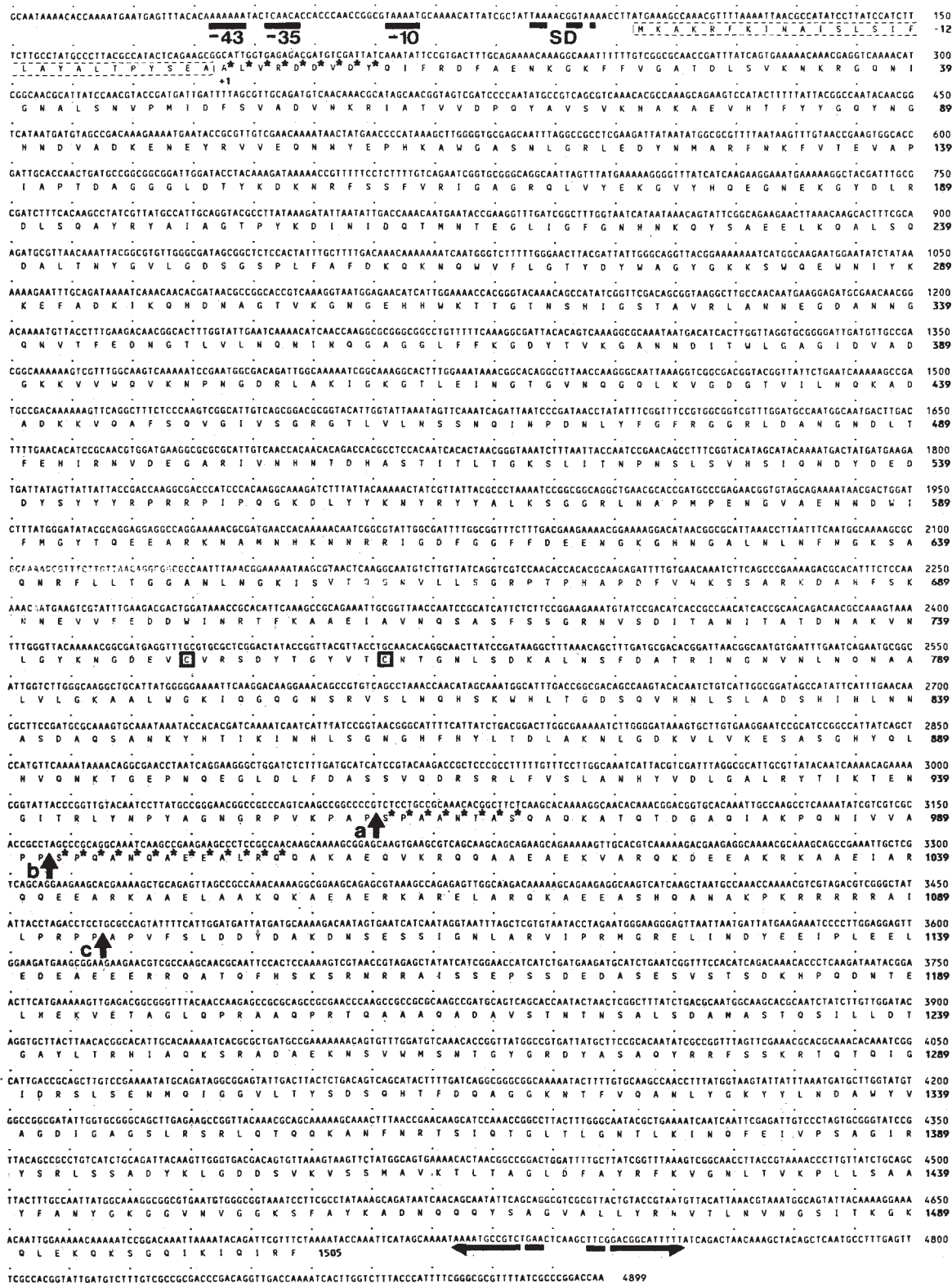
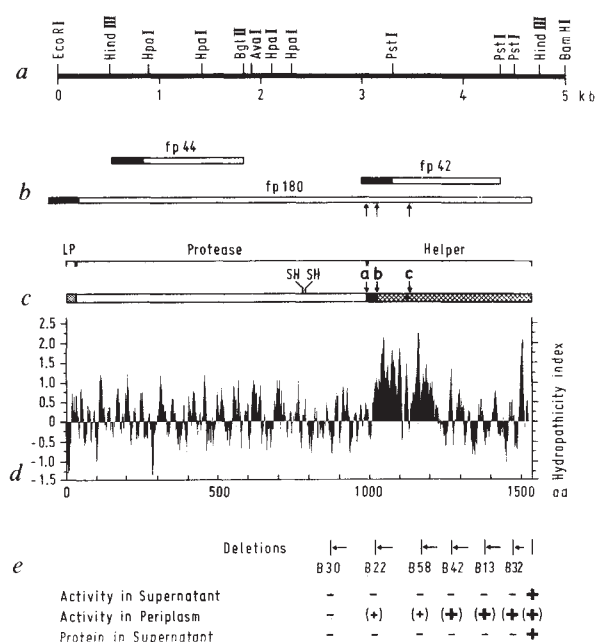


Fig. 1 Sequence of the *N. gonorrhoeae* MS11 *iga* gene. The DNA sequence reveals a single gene of 4,596 bp in association with typical expression signals: a promoter sequence¹⁷⁻¹⁹ (positions -43, -35, and -10), a transcription termination signal²⁰ (divergent arrows), and a Shine-Dalgarno (SD) sequence²¹ combined with an ATG start codon. The amino-acid sequence is a 169 K precursor of 1,532 amino acids. *, Coincidence of the DNA sequence with the amino-terminal protein sequences of purified protease and the autoproteolytic cleavage sites of IgA protease in its precursor; broken frame, the amino-terminal leader peptide of the precursor; solid frame, the two unique cysteines possibly associated with the active centre. Positions a, b and c represent autoproteolytic sites for IgA protease. Essential features of the precursor primary structure are summarized in Fig. 2.

Methods. Cloning of the *N. gonorrhoeae* MS11 IgA protease gene and construction of pIP100 has been described⁴. The nucleotide sequence is the sum of overlapping sequences of both strands of the pIP100 insert. Maxam and Gilbert sequencing²² was performed on 5'-labelled restriction fragments of the pIP100 insert and suitable *Bal*31 deletion derivatives (Fig. 2 legend). Restriction fragments from pIP100 were also subcloned in M13mp8 and mp9 phages and sequenced by the dideoxy method²³. Amino-terminal protein sequence analysis was performed on a mixture of the 106 and 109 K forms of purified IgA protease and with the 60 K products from cleavage of fp180 with IgA protease (Fig. 4a). Both proteins (150 pmol each) were extracted from gels, lyophilized, dissolved in 60 μ l formic acid, and two sequencing runs were performed with 75 pmol loadings using an ABS-A470 protein sequencer with on-line HPLC analysis. The two IgA protease forms present in the mixed sample had identical amino termini suggesting carboxy-terminal heterogeneity corresponding to sites a and b. Sequencing of the 60 K cleavage product gave double signals in steps 3, 6 and 8, clear enough to determine sites a and b as two separate amino termini formed by cleavage (Fig. 4a) of fp180 with IgA protease.

Fig. 2 Structural features of gonococcal protease and its gene. *a*, Physical map of the *iga* gene in pIP100⁴. *b*, Fusions of the *iga* gene with the pEX31B expression vector. This vector elicits fusion proteins (fp) with a 12 K amino-terminal fraction (solid lines) of the bacteriophage MS2 polymerase. *c*, Linear map of the IgA protease precursor. The existence of a leader (LP, dashed section) was deduced from the DNA sequence and from the absence of a corresponding protein segment in purified protease. The protease domain (open section) was located in the precursor by amino-terminal sequence analysis of purified protease, and by immunological mapping with fusion proteins. Antisera raised against purified IgA protease cross-reacted with fp44 and fp180 but not with fp42; anti-fp44 and anti-fp180 reacted with purified IgA protease but there was no cross-reaction with anti-fp42. The α - and β -domains of the helper (crossed section) are defined by their structural properties according to Chou and Fasman²⁴ and by their locations downstream of the internal cleavage sites a and c, respectively. Arrows, internal IgA protease target sites (a, b and c). SH, positions of two unique cysteines in the precursor. *d*, the hydrophathy plot of the precursor primary structure according to Hopp and Woods²⁵, using a window of 11 consecutive residues. *e*, Extent of 3'-terminal deletions of the *iga* gene and their effects on the concentration and activity of IgA protease in the culture supernatant and periplasm of *E. coli* GCl host cells.

Methods. *b*, To produce *iga* fusion proteins, the *iga* gene in pIP100 was digested at its 5'-terminal *Eco*RI site using *Bal*31 exonuclease⁴. The deleted genes were fused with *Eco*RI linkers and placed under control of the λ P₁ promoter in the pEX31B vector⁹. Several clones were found to produce larger *iga* fusion proteins of up to 180 K (Fig. 4a). One plasmid, pFP180, was used to prepare pFP44 by deleting the 3'-terminal *Bgl*II/*Bam*HI fragment of the *iga* gene. Clone pFP42 was obtained by inserting the 1-kb *Pst*I fragment from pIP100 into the polylinker region of pEX31B. *e*, The 3'-terminal deletions in the *iga* gene were generated by *Bal*31 exonuclease digestion initiated at the *Bam*HI site of pIP100. The *Bam*HI sites were reconstituted with synthetic linkers and the truncated *iga* genes inserted into pBR322 vector. Deletions were characterized by restriction mapping and DNA sequencing²². Extracellular *iga*-encoded protein was assayed by immunoblotting (see Fig. 4a for clone pIP-B32). Extracellular and periplasmic activity of IgA protease was determined by fractionation of *E. coli* clones (legend to Fig. 3).



a carboxyl-terminal 'helper' domain apparently required for extracellular secretion (excretion). Based on the structural features of the precursor, we propose a model in which the helper serves as a pore for excretion of the protease domain through the outer membrane. IgA protease acquires an active conformation as its extracellular transport proceeds and is released as a proform from the membrane-bound helper by autoproteolysis. The soluble proform further matures into the 106 K IgA protease and a small stable α -protein.

The nucleotide sequence of the cloned *iga* locus in plasmid pIP100 (ref. 4), shown in Fig. 1, contains a large open reading frame which codes for a precursor of IgA protease. The predicted M_r of this precursor is ~169 K exceeding the size of the mature extracellular protease by 63 K (Fig. 2c). Using the pEX expression system⁹, we confirmed the existence of the precursor. We obtained fusion proteins of up to 180 K with the *iga* gene (Figs 2b, 4a).

That the mature extracellular IgA protease maps towards the amino terminus of the precursor can be concluded from an analysis of the immunological cross-reactivity between various fusion proteins and purified IgA protease (Fig. 2 legend). Also, amino-terminal sequence analysis of purified enzyme (Fig. 1) locates the amino terminus of extracellular protease to amino acid position 28 in the precursor, suggesting an amino-terminal signal peptide of 27 amino acids. The sequence in this segment is typical of a prokaryotic signal peptide¹⁰ showing four positive amino-terminal charges, a central hydrophobic area, and an Ala-Ala signal peptidase cleavage site (Fig. 1). Taking into account the M_r of 106 K for mature IgA protease and its relative location we can assign three domains in the precursor of IgA protease, the amino-terminal leader, the actual protease domain, and a carboxy-terminal region (Fig. 2c). This carboxy terminal region was named the helper because it turned out to be essential for the excretion of IgA protease.

Gonococcal IgA protease is excreted not only from the original host system but also from hybrid *E. coli* containing the

iga gene⁴. To ensure that the release of protease from this foreign host is not due to a general leakiness of the outer membrane, the distribution of IgA protease and appropriate marker enzymes in different compartments of *E. coli* cells was determined (Fig. 3). The experiment clearly indicates that both cytoplasmic and periplasmic marker enzymes are retained in the *E. coli* host, regardless of the production and excretion of IgA protease.

To determine the role of the helper in excretion of IgA protease, we constructed a series of mutants with 3'-terminal deletions of this region and tested them for their potential to excrete active or inactive enzyme into the periplasm or the extracellular environment of the *E. coli* host. Even with the smallest deletion derivative, we observed complete loss of activity in the culture supernatant (Fig. 2e). No IgA protease specific protein was detectable by immunoblotting in culture supernatants of mutant derivatives, whereas the *E. coli* (pIP100) positive control clearly showed the 106 K band of extracellular IgA protease (Fig. 4b). This suggests an essential role for the helper in the outer membrane transport of IgA protease. However low levels of IgA protease activity, not exceeding the activity in wild-type controls, were found in the periplasm with the mutant *iga* genes (Fig. 2e) suggesting that the precursor cannot be efficiently converted into an active enzymatic form in the periplasm.

The amino-acid sequence of the precursor contains only two cysteines. They are located in the protease domain and only 11 amino acids apart. Low cysteine content is characteristic of other secretory proteins¹¹. The two cysteines are presumably part of the active centre of IgA protease as the enzyme is sensitive to treatment with heavy metals (unpublished data).

The amino-terminal region of the helper appears as a strongly hydrophilic area in the hydrophathy profile (Fig. 2d) and contains proline-rich sequences (a, b and c in Figs 1, 2c), that are reminiscent of the cleavage site of gonococcal IgA protease in human IgA₁ (ref. 12). To test if these sites are potential targets for IgA protease, we used IgA protease fusion protein fp180

Fig. 3 Distribution of IgA protease and marker enzymes in *E. coli* GCl cultures. Numbers on top of the blocks indicate the percentage of total enzyme activity. IgA protease and marker enzymes, β -lactamase (periplasmic), cyclic phosphodiesterase (periplasmic), and β -galactosidase (cytoplasmic), were tested in the culture supernatant, in the soluble fraction of shocked cell suspension (periplasm), and within the shocked cells (cytoplasm).

Methods. Cells were grown to the late logarithmic phase and subjected to osmotic shock²⁶. The activities of IgA-protease⁴, β -galactosidase²⁶, and cyclic phosphodiesterase²⁷ in the fractions were determined using standard assays. PADAC (Calbiochem) was used as substrate to assay β -lactamase²⁸.

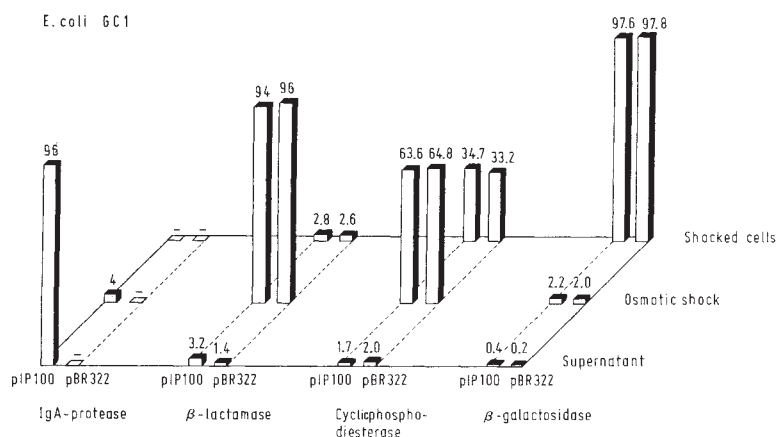


Fig. 4 Analysis of IgA protease, its autologous substrates and its secretory intermediates, by gel electrophoresis and immunoblotting.

a. Determination of autoproteolytic sites in the IgA protease precursor. Lanes 1–3, Coomassie-stained gel; 1, fusion protein fp180 (compare with Fig. 2b); 2, fp180 after cleavage with IgA protease, yields major (120 and 60 K) and minor (45 and 15 K) products; 3, purified IgA protease. Lanes 4–6, same as lanes 1–3, but immunoblotted with monoclonal antibody HAL-1 which is directed against mature IgA protease and gives positive reaction with fp180, the 120 K product and mature protease. Lanes 7–9, same as lanes 1–3, but immunoblotted with anti-fp42 showing positive reaction with the 60, 45 and 15 K products; reaction with the 120 K product is due to the MS2 portion in fusion proteins. **b.** Secretory intermediates of IgA protease. Lanes 10–18, immunoblot of crude concentrated culture supernatants using HAL-1 antibody. Culture supernatants of *E. coli* GCl containing plasmid pIP100 (lanes 10, 12, 14 and 16) and the deletion derivative pIP-B32 shown in Fig. 2e (lanes 11, 13, 15 and 17) were harvested after different periods of growth at an OD₆₅₀ of 0.55 (lanes 10 and 11), 1.3 (lanes 12 and 13), 1.6 (lanes 14 and 15), and 2.0 (lanes 16 and 17). Lane 18, immunoblot of an *N. gonorrhoeae* MS11 supernatant using HAL-1 antibody. **c.** Localization of the helper in the outer membrane. Lanes 19 and 20, immunoblot with HAL-2, a monoclonal specific for the β -domain; 19, fp180 after treatment with IgA protease; 20, outer membrane preparation²⁹ of *N. gonorrhoeae* MS11. Lanes 21–25, immunoblot of the soluble α -protein with anti-fp42 serum; 21, same sample as in lane 19; (lanes 22, 23), crude supernatant from *N. gonorrhoeae* MS11 cultures and (lanes 24, 25) from two *E. coli* (pIP100) cultures.

Methods. Active IgA protease was prepared from culture supernatants as described⁴. Fusion proteins were purified by a modification of the method of Hunkapiller *et al.*³⁰ and eventually digested with IgA protease using an enzyme/substrate ratio of 1:50 in 10 mM Potassium phosphate pH 7.5, 150 mM NaCl (60 min at 37 °C). Proteins in crude culture supernatants were precipitated with 10% trichloroacetic acid on ice and collected by centrifugation. Samples were separated on 14% (a) and 12.5% (b, c) SDS polyacrylamide electrophoresis gels.

(Fig. 2b) as substrate for cleavage with IgA protease in *trans*. This reaction resulted in specific degradation products which were further identified by immunoblotting (Fig. 4a). Two of the products were subjected to amino-terminal sequence analysis (Fig. 1). From these data, we conclude that sites a and b are efficient targets for IgA protease giving rise to products of M_r ~120 K and 60 K. The minor 45 K and 15 K products result from partial cleavage of the 60 K product at site c (Figs 2b and 4a).

We have repeatedly observed an IgA protease-specific double band of 106/109 K (previously estimated as 104/105 K) in the purified IgA protease samples⁴ which can be explained as extracellular enzymes with different carboxy-terminal ends corresponding to sites a and b. In crude culture supernatants another protein related to IgA protease can be detected, the carboxy terminus of which matches site c (Figs 2c and 4b). This protein of 121 K not only cross-reacts with anti-protease serum (Fig. 4b) but also with an antiserum raised against fp42 (not shown). We have determined the relative amounts of the three

forms of extracellular IgA protease in a growing *E. coli* culture (Fig. 4b). In the early growth phase, the 121 K intermediate of the IgA protease is prominent. As the enzyme accumulates, the 109 K intermediate becomes the major band and finally the mature 106 K form of IgA protease is clearly detectable. This suggests that the high molecular weight intermediates are slowly converted into the mature enzyme. The conversion can also be seen with purified enzyme fractions (data not shown).

The detection of an extracellular 121 K proform implies that the strongly polar, α -helical part of the helper (α -protein, Fig. 2c, d) is excreted, along with IgA protease, and subsequently cut off. This was shown directly by immunoblotting using anti-fp42 serum to screen culture supernatants for the α -protein enclosed within target sites a/c or b/c (Fig. 2c). The α -proteins can be seen for both gonococci and *E. coli* (pIP100) (Fig. 4c).

The carboxy terminal portion of the helper can form amphipathic β -sheet structures suggesting membrane association (Jähnig, F., unpublished data). To test this we prepared outer mem-

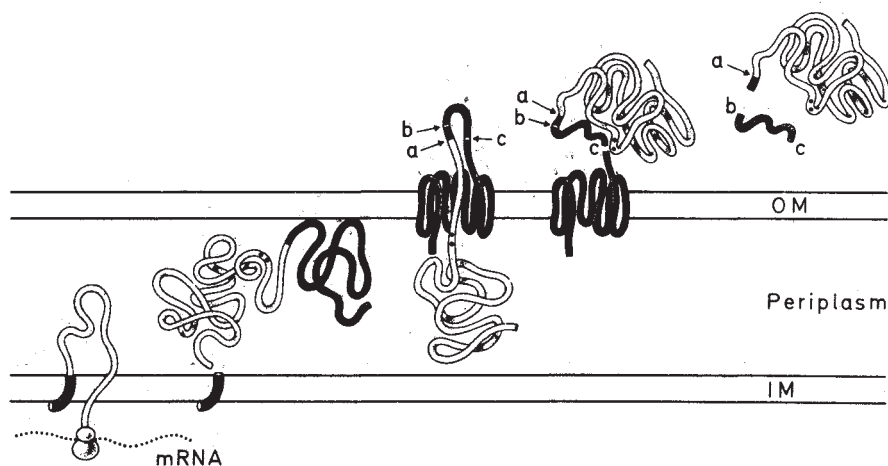


Fig. 5 Model for extracellular secretion of IgA protease. The leader peptide guides the IgA protease precursor into the periplasmic space via the regular co- or post-translational transport route taken by the majority of periplasmic and outer membrane proteins (ref. 31 for review). In a second step, the carboxy-terminal helper associates with the outer membrane to form a pore for excretion of the protease domain. This domain, still connected to the helper, gains its active conformation in the process of secretion. A proform of the IgA protease is released from the helper by autoproteolysis and further develops into α -protein and the mature enzyme.

brane fractions from gonococci and analysed them for the presence of the β -protein. Using a β -domain specific monoclonal antibody raised against fp42, we detected in the outer membrane of gonococci a 45 K protein (Fig. 4c), which precisely matches the β -domain in size.

Based on these results, we propose a model for the extracellular secretion of IgA protease (Fig. 5). We assume that the signal peptide allows the precursor to pass the inner membrane. Owing to its amphipathic features the helper becomes incorporated into the outer membrane to form a pore for the excretion of the protease domain and the α -protein. During excretion through the helper pore, IgA protease acquires an active conformation and is released from the helper by autoproteolysis. The extracellular form of IgA protease further develops by autoproteolysis to yield mature protease and the α -protein. This two-step mechanism is supported by the detection of residual IgA protease activity in the periplasm of cells harbouring either an intact *iga* gene or a helper mutant gene. A definitive distinction however is not possible between this model and a one-step secretion mechanism, using the inner membrane transport machinery, where the helper function is to direct the enzyme to a putative inner/outer membrane junction^{13,14}.

Few other examples of protein excretion in Gram-negative bacteria are known in detail¹⁵⁻¹⁷. The gonococcal IgA protease system is striking in that a single protein serves all specific functions of its own secretion, not only in the authentic host but also in the foreign environment of *E. coli*. It may prove a suitable system for manipulating and studying protein export in Gram-negative organisms. With respect to the virulence function of gonococcal IgA protease, our demonstration that IgA₁ is not an exclusive target of this enzyme suggests that other sensitive targets might exist in humans. Two other components, the membrane-associated helper and the soluble α -protein, appear to be intrinsically connected with IgA protease, it might be worthwhile to consider the significance of these factors in gonococcal pathogenesis.

We thank K. Neifer and B. Tries for technical assistance, K. Lamberti for artwork, R. Haas for computer assistance, G. Jähnig, J. R. McGhee and K. Muralidharan for critical suggestions, E. Beck and H. Schaller for releasing the pEX vector system before publication, S. Schweitzer and R. Sommer of Genbiotech GmbH for collaborating on the DNA analysis, and M. Ganss for hybridoma work. The project was supported through BMFT grant BCT 0364-1 and through generous commitments of H. Hoffmann-Berling and K. Rother.

Received 3 July; accepted 10 November 1986.

1. Plaut, A. G., Gilbert, J. V., Arstenstein, M. S. & Capra, J. D. *Science* **1290**, 1103-1105 (1975).
2. Kilian, M., Mestecky, J. & Schrohenloher, R. E. *Infect. Immun.* **26**, 143-149 (1979).
3. Koomey, J. M., Gill, R. E. & Falkow, S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7881-7885 (1982).
4. Halter, R., Pohlner, J. & Meyer, T. F. *EMBO J.* **3**, 1595-1601 (1984).

5. Bricker, J., Mulks, M. H., Plaut, A. G., Moxon, E. R. & Wright, A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2681-2685 (1983).
6. Fishman, Y., Bricker, J., Gilbert, J. V., Plaut, A. G. & Wright, A. in *Pathogenic Neisseriae Proc. Fourth Int. Microbiol.* (eds Schoolnik, G. K. et al.) 164-168 (1985).
7. Rahr, S., Halter, R., Müller, H., Pohlner, J. & Meyer, T. F. in *Pathogenic Neisseriae Proc. Fourth Int. Symp. Soc. Microbiol.* (eds Schoolnik, G. K. et al.) 157-163 (1985).
8. Cooper, M. D., McGee, Z. A., Mulks, M. H., Koomey, J. M. & Hindman, T. L. *J. inf. Dis.* **150**, 737-744 (1984).
9. Strebel, K., Beck, E., Strohmaier, K. & Schaller, H. *Virology* **57**, 983-991 (1986).
10. Watson, M. E. E. *Nucleic Acids Res.* **12**, 5145-5164 (1984).
11. Pollock, M. R. & Richmond, M. H. *Nature* **194**, 446-449 (1962).
12. Mulks, M. H., Plaut, A. G., Feldman, H. A. & Frangione, B. *J. exp. Med.* **152**, 1442-1447 (1980).
13. Lory, S., Tai, P. C. & Dabis, B. D. *J. Bact.* **156**, 695-702 (1983).
14. Bayer, M. E. in *Bacterial outer membranes: biogenesis and function* (ed. Inouye, M.) 167-202 (Wiley, New York, 1979).
15. Wagner, W., Vogel, M. & Goebel, W. *J. Bact.* **154**, 200-210 (1983).
16. Mackman, N., Nicaud, J.-M., Gray, L. & Holland, I. B. *Molec. gen. Genet.* **201**, 529-536 (1985).
17. Hawley, D. K. & McClure, W. R. *Nucleic Acids Res.* **11**, 2237-2255 (1983).
18. Schaller, H., Grey, C. & Herrmann, K. *Proc. natn. Acad. Sci. U.S.A.* **72**, 737-745 (1975).
19. Gentz, R. Bujard, H. *J. Bact.* **164**, 70-77 (1985).
20. Rosenberg, M. & Court, D. A. *Rev. Genet.* **13**, 319-353 (1979).
21. Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1342-1346 (1974).
22. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
23. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
24. Ghosh, P. Y. & Fasman, G. D. A. *Rev. Biochem.* **47**, 251-276 (1978).
25. Hopp, T. P. & Woods, K. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3824-3828 (1981).
26. Neu, H. D. & Heppel, L. A. *J. biol. Chem.* **240**, 3685-3692 (1965).
27. Craven, G. R., Steers, E., Jr & Anfinsen, C. B. *J. biol. Chem.* **240**, 2468-2473 (1965).
28. Howard, S. P. & Buckley, J. T. *J. Bact.* **161**, 1118-1124 (1985).
29. Stern, A., Brown, M., Nickel, P. & Meyer, T. F. *Cell* **47**, 61-71 (1986).
30. Hunkapiller, M. W., Lujan, E., Ostrander, F. & Hood, L. E. *Meth. Enzym.* **91**, 227-236 (1983).
31. Randall, L. L. & Hardy, S. J. S. *Microbiol. Rev.* **48**, 290-298 (1984).

Erratum

Energy conversion catalysis using semiconducting transition metal cluster compounds

N. Alonso Vante & H. Tributsch

Nature **323**, 431-432 (1986).

IN this letter, the metal atoms in the Fig. 1 legend were incorrectly defined. They should read: ●, Mo; ⊖, Ru; ○, Se.

US Supreme Court to review Louisiana appeal

Thomas H. Jukes

Nature **324**, 423-424 (1986).

IN this commentary a sentence in the seventh paragraph contained an incorrect age for the Earth and the Universe. It is correct as printed here.

"This places the date of creation as about 10,000 years ago, which accounts for opposition by creationists towards any of the natural sciences that state that the Earth and the Universe are thousands of millions of years old."

Observation of an unexplained event from a magnetic monopole detector

IN a recent article Caplin *et al.*¹ reported an event, event-160. They rejected many possible terrestrial causes for event-160 and offered the event as a candidate magnetic monopole sighting. Here I demonstrate that although the precise cause of event-160 remains unknown the experimental record of the event contains enough evidence to discount any connection with a magnetic monopole.

The design of this experiment intentionally sacrificed uniqueness of signal size and full coincidence for maximum detection area. As a result the only check against fast, spurious single-channel events was the simultaneous records of four sensors that monitored the mechanical and electromagnetic environment. The experimental method was to reject any event whose record contained a real anomaly in just one of these monitors even though a definite causal connection between monitor feature and SQUID offset was not always found². The record event 160 contains a complicated and rare anomaly in the radiofrequency (r.f.) monitor. In this experiment the choice between artefact and serious candidate hinges on our understanding of this monitor.

Caplin *et al.* correctly pointed out that the r.f. signal was dominated by the stray radiation associated with data transmissions to and from the computer. They suggested that the r.f. anomaly was a direct consequence of the change in sign in the numbers transmitted by the SQUID event channel. A careful analysis of the data shows that the r.f. anomaly was composed of three parts, two of which clearly precede the SQUID offset. About 40 s before the event the mean signal level began a gradual rise. About 0.4 s before the offset the variance of the signal dropped abruptly to its lowest level in the entire record. At this instant the mean level began a more rapid rise that persisted across the time of the SQUID offset. Caplin *et al.* restricted their attention to this last component.

A statistical analysis of 40 preserved records each consisting of 8000×7 data points was made. Only two other records were found to contain the features seen in event-160. These were readily explained and were not associated with any sign change. The same survey showed that sign changes in at least one channel were the norm and that there was no evidence for any associated r.f. feature.

In a long series of experimental tests square-wave input signals were used to simulate large, fast offsets in all three SQUID channels. The system was subjected to >100 trials at different times of day and with different test amplitudes. Of these

only one batch of ~15 trials produced any response at all on r.f. In this batch the effect had the wrong sign to explain event-160: it was clearly visible on other channels and was subsequently traced to amplifier saturation caused by an excessively large input signal. The magnitude of event-160 was two orders of magnitude too small to saturate its amplifier.

A later investigation of the r.f. channel itself revealed that amplifier slew rate and pulse distortion made it impossible for r.f. to contain any information about any bit pattern being transmitted by any channel. The same study provided enough information to permit the reproduction at will of the main features of variance and mean shift seen in event-160 using changes in computer speed and antenna coupling. These last tests showed that quite significant changes were required to produce any effect on r.f.

It is very unlikely that two very rare features should occur within 0.4 s in the same record by chance. I have shown that the r.f. anomaly preceded the offset. I have shown that the r.f. anomaly could not have been generated without a significant change either in the computer or the environment of the apparatus. I therefore conclude that such a change was responsible for both recorded features.

C. N. GUY

Blackett Laboratory,
Imperial College,
London SW7 2BZ, UK

1. Caplin, A. D., Hardiman, M., Koratzinos, M. & Schouten, J. *Nature* **321**, 402–406 (1986).
2. Caplin, A. D., Guy, C. N., Hardiman, M., Park, J. G. & Schouten, J. *Nature* **317**, 234–236 (1985).

CAPLIN *ET AL.* REPLY—The r.f. feature alluded to by Dr Guy in connection with event-160 is not unique; indeed, during the year of normal operation of the detector, several similar features were seen.

Most of the 170 offsets recorded by our detector¹ were small, but the mechanism probably responsible for these r.f. features (see below) is likely to be operative only with fast offsets that are larger than peak-to-peak noise. Further, external noisy r.f. sources could obscure the effect. The susceptible category of offsets is therefore a small fraction of the total.

Of the 40 putative events surveyed by Guy, which are those for which the full 20-Hz data on floppy disk have been retained, only 5 (event-0 caused by mechanical shock, events-47, 50 and 75 caused by an intermittent electronic fault in a SQUID control unit, and the unexplained event-160) fall into the susceptible category, apart from a few offsets that were so large as to saturate amplifiers in the data acquisition system. Events-47 and 50 show r.f. shifts of 2 and 0.7 V respectively, which should be compared with the 0.1 V r.f. shift associated with event-160 (the

mean value of the r.f. level is typically 5 V, with fluctuations of 2 V peak-to-peak).

Of the other 130 putative events, the causes were sufficiently well identified and lacking in interest that the floppy disks were erased; even so, some sections of the detailed record were usually retained in printed form. Susceptible events for which relevant printed records are available are events-1 (mechanical shock) and 84, 85 and 90 (intermittent electronic fault in a SQUID control unit). Events-1 and 90 show r.f. shifts of about 0.5 and 0.3 V respectively.

Thus the experimental evidence is that in normal operation, about half of the large and sudden offsets in a detector channel are accompanied by features in the r.f. record. The computer-controlled data acquisition system provides the likely mechanism for this coupling¹; under quiet conditions it is responsible for 90% of the recorded r.f. interference. A large change in one of the input channels can therefore affect the radiated r.f., through an interaction at the hardware level, through a perturbation of the rhythm of program execution, or both.

We do not understand in detail why, even under the more or less constant conditions of normal operation, the form and amplitude of the r.f. feature should be so variable; perhaps the computer speed and antenna coupling referred to by Guy are important. The attempted simulations that he describes were made after the cryostat had been warmed to room temperature and the detector dismantled; the results obtained under those very different conditions are of limited relevance to behaviour in normal operation.

Guy's conclusion that the r.f. feature and event-160 were generated by a (unidentified) 'significant change either in the computer or the environment of the apparatus' not only fails to identify a causative mechanism, but ignores the ample experimental evidence showing that the r.f. feature is likely to be a mere artefact of the event.

Other mechanisms that could have generated event-160, and which are physically reasonable, were discussed in detail in our article¹.

A. D. CAPLIN*
M. HARDIMAN†
J. C. SCHOUTEN‡

*Physics Department,
Imperial College,
London SW7 2BZ, UK

†School of Mathematical and
Physical Sciences,

University of Sussex,
Brighton BN1 9QH, UK

‡Oxford Instruments Ltd,
Osney Mead,
Oxford OX2 0DX, UK

1. Caplin, A. D., Hardiman, M., Koratzinos, M. & Schouten, J. C. *Nature* **321**, 402–406 (1986).

Gel electrophoresis of protein-DNA complexes

D.M. Crothers

Regions of DNA bending can be explored using gel electrophoresis on protein-DNA complexes, proving gel electrophoresis does more than just separate proteins and nucleic acids according to size.

GEL electrophoresis is among the most commonly used techniques in molecular biology, where it has virtually replaced the use of older physical methods for determination of molecular weight. The extraordinary resolving power of the gel sieving effect is illustrated by the ability of DNA sequencing gels to separate polymeric molecules differing by only one residue. This resolution far surpasses that of ultracentrifugation, or even that of high resolution chromatographic techniques.

The advantages of classical hydrodynamic methods, such as centrifugation, have been their ability to address problems that go beyond simple separations, to issues of conformational changes and macromolecular interactions. Recent work on the problem of specific protein-DNA interactions and the phenomenon of DNA bending demonstrates that gel electrophoresis has great power when applied to these questions. Intrinsic advantages of the method include high resolution and the ability to analyse tiny amounts of material. These combine to offer remarkable potential for the characterization of interaction specificity, even in complex systems, and provide a novel and scarcely explored approach to the intriguing problem of detecting variations in DNA and protein shape. The lack of a fully quantitative theory for interpreting electrophoretic mobilities, however, has hampered these efforts.

Protein-DNA complexes

It has been known for some years that tightly bound protein-nucleic acid complexes are stable during electrophoresis under non-denaturing conditions, as found, for example, in RNA polymerase-DNA complexes¹, and in nucleosomes with or without bound HMG protein². Intensive application of non-denaturing gel electrophoresis to regulatory protein-DNA complexes began in 1981 with experiments by Garner and Revzin³ on *Escherichia coli* CAP protein and by Fried and Crothers⁴ on *lac* repressor, which demonstrated that the DNA complexes could be detected and characterized quantitatively by the gel method.

Figure 1 illustrates the advantages that the gel method has over the standard filter binding technique for detecting protein-DNA complexes. When *lac* repressor is added in increasing amounts to *lac* promoter DNA, a series of gel bands is observed (lanes b-j), of regularly decreasing mobility, which contain increasing numbers of

bound repressor molecules⁴. Hence the products of the protein-DNA interaction are separated into discrete species, rather than grouped together in the "filter bound" category. The low-salt conditions in the gel, and a putative "cage effect"¹⁴ created by the gel matrix, stabilize even weak complexes so that they can be quantified, enabling determination of binding equilibria and kinetics³⁻⁷. As recently shown by Liu-Johnson *et al.*⁸, one can even dissect the binding affinity into contributions by individual base pairs in the binding site.

As might be expected, protein size affects the gel mobility of the complex.

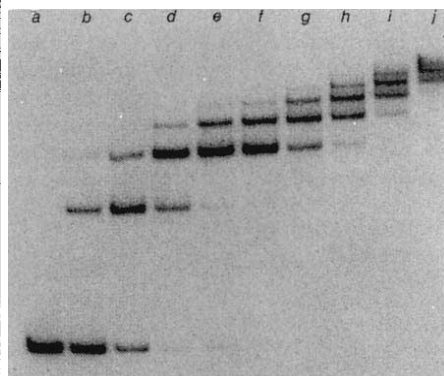


Fig. 1 Titration of free *lac* promoter DNA (lane a) with increasing amounts of *lac* repressor protein (lanes b-j). Increasing the number of repressor molecules bound yields bands (detected by ³²P autoradiography) of regularly decreasing mobility.

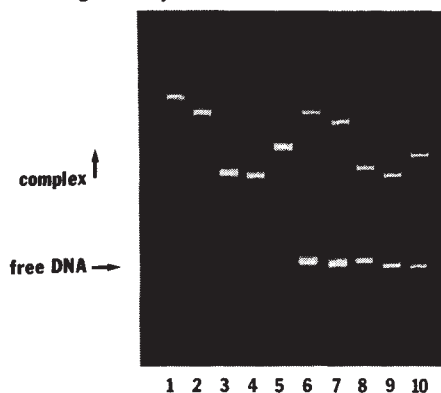


Fig. 2 Dependence of the mobility of CAP protein-*lac* promoter DNA complexes (detected by ethidium fluorescence) on circular permutation of the DNA sequence. Lanes 1-5 correspond to the wild type CAP binding sequence, and lanes 6-10 to a mutant site symmetrized on the half of the CAP site that is closer to the promoter. In the slowest moving bands (lanes 1 and 6) the protein binding site and its induced bend are located near the middle of the DNA molecule⁸.

Hendrickson and Schleif⁹ exploited this feature in examining the interaction of the arabinose promoter with a mixture of several proteins. Each complex produced a distinctive band. In addition, the conformation of a protein influences its mobility on the gel, as shown by the different mobilities of RNA polymerase-promoter "open" complexes observed by Straney and Crothers⁹. Probably the most widespread current use of this "band-shift" or gel retardation method is in the assay of complex nuclear protein mixtures for their ability to produce a discrete protein-DNA gel band due to binding to a specific DNA sequence¹⁰.

DNA bending

It is now generally agreed that DNA structure is pliable, and can even reverse its handedness from the classic right-handed B-DNA double helix to form left-handed Z-DNA. Another striking structural polymorphism in which DNA is intrinsically bent rather than straight was discovered as a result of its anomalous gel electrophoretic properties¹¹. DNA molecules from the kinetoplast body of parasites such as *Leishmania* contain bending sites consisting of tracts of 5-6 homopolymeric A-T pairs repeated at 10 base-pair intervals. Because their spacing is in phase with the DNA helix screw, the bends add to produce a pronounced overall curvature of the molecule. These bent helices migrate more slowly than straight helices in electrophoretic gels because their average end-to-end distance, on which theory indicates the gel mobility should depend^{12,13}, is smaller than for normal DNA.

As Wu and Crothers¹⁴ showed, the location of a bend in a DNA sequence can be found by comparing the electrophoretic mobilities of a series of circularly permuted variants of the same sequence. Isoomers with the bend near the end have a much larger end-to-end distance, and hence greater gel mobility, than that observed for molecules in which the bend is near the middle of the chain.

Demonstration of the distinctive electrophoretic manifestation of intrinsic DNA bending has made it possible to detect DNA bending induced by protein binding¹⁴. Figure 2 shows the variation in gel mobility for two series of circularly permuted DNA molecules containing bound *E. coli* CAP protein, attached either to the wild type *lac* promoter sequence (lanes 1-5), or to a chemically synthesized symmetric variant (lanes 6-

10)⁸. The slowest moving species (lanes 1 and 6) have the bound protein located near the centre of the molecule. Control experiments using *lac* repressor-DNA complexes do not demonstrate variable mobility on electrophoretic gels, no matter where the protein is bound. It is therefore concluded that the variation of mobility with CAP-DNA complexes is due to DNA bending induced by the protein¹⁴. It may be of general functional significance that the gene activating protein induces a DNA bend whereas the repressor protein does not.

Comparison of lanes 1 and 6 shows that mobility is slower for the wild type binding sequence than for the symmetrized variant, indicating that bending is greater in the wild type sequence⁸. However, comparison of lanes 3 and 8, in which the protein binding site is very near the end of the DNA, shows that mobility in this case is greater for the wild-type sequence (lane 3). Lui-Johnson *et al.*⁸ argue that this observation suggests that the induced bend is between 90° and 180°, because the end-to-end distances of the species in lanes 3 and 8 are very similar, and therefore the frictional effects are dominated by the protein and the small piece of DNA that protrudes from one side. For a bend in the 90–180° range, the decreasing extent of bending in the symmetrical variant (lane 8) causes the effective width of the protein and protruding DNA to increase, thus increasing the frictional resistance and decreasing the mobility.

Given the lack of a quantitative theory of gel electrophoretic mobilities, such interpretations should be viewed with a measure of healthy scepticism. However, as the empirical evidence mounts, it should be possible to "bootstrap" a consistent set of rules for interpreting gel mobilities of protein-DNA complexes, thus providing a powerful general tool for characterizing the topology of DNA in protein-bound complexes. □

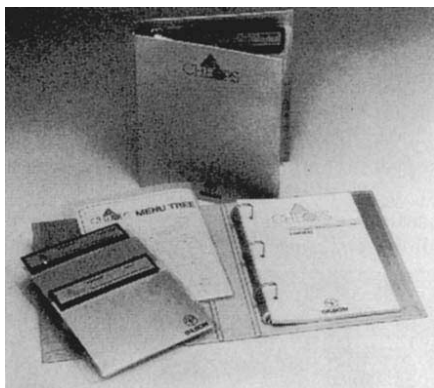
Donald M. Crothers is at the Department of Chemistry, Yale University, New Haven, Connecticut 06511, USA.

- Chelm, B.K. & Geidushek, E.P. *Nucleic Acids Res.* **7**, 1351–1367 (1975).
- Sandeel, G., Wood, W.I. & Felsenfeld, G. *Nucleic Acids Res.* **8**, 3757–3778 (1980).
- Garner, M.M. & Revzin, A. *Nucleic Acids Res.* **9**, 3047–3060 (1981).
- Fried, M.G. & Crothers, D.M. *Nucleic Acids Res.* **9**, 6505–6525 (1981).
- Fried, M.G. & Crothers, D.M. *J. molec. Biol.* **172**, 241–262 (1984).
- Fried, M.G. & Crothers, D.M. *J. molec. Biol.* **172**, 263–282 (1984).
- Hendrickson, W.F. & Schleif, R.F. *J. molec. Biol.* **178**, 611–628 (1984).
- Liu-Johnson, H.-N. Gartenberg, M.R., & Crothers D.M. *Cell* **47**, 995–1005 (1986).
- Straney, D.C. & Crothers, D.M. *Cell* **43**, 449–459 (1985).
- Carthew, R.W., Chodosh, L.A., & Sharp, P.A. *Cell* **43**, 439–448 (1985).
- Marini, J.C., Levene, S.D., Crothers, D.M., & Englund, P.T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7664–7668 (1982).
- Lerman, L. & Frisch, H.L. *Biopolymers* **21**, 995–997 (1982).
- Lumpkin, O.J. & Zimm, B.H. *Biopolymers* **21**, 2315–2316 (1982).
- Wu, H.M. & Crothers, D.M. *Nature* **308**, 509–513 (1984).

Separation showcase

Chromatography and electrophoresis are the themes for this week's issue, with something for every separation need.

FOR the optimization of the mobile phase in gradient elution HPLC. Gilson Medical Electronics offers its new Chromatographic Elution Optimization Software, CHEOPS (Reader Service No. 100). The



Makes the most of your mobile phase.

CHEOPS program is compatible with any chromatographic system, and runs on Apple II microcomputers equipped with two disk drives, a video monitor and a printer-plotter. CHEOPS generates simulated chromatograms with expected retention times for each compound, using a choice of two gradients: polarity alone, with up to four solvents, or polarity molarity and pH. Optimization of the separation of up to 19 peaks is possible with CHEOPS, says Gilson, with desired analysis time, maximum analysis time and resolution weighting factors selected by the user. The CHEOPS system comes with a method diskette, a tutorial diskette containing practice data files, and a complete user's guide.

Trivector is selling a way to network laboratory gas chromatography. Its TRIO Chromatography Integrator has two channels, allowing it to process data from up to two gas chromatographs, and store data permanently on a disc unit for reanalysis (Reader Service No. 101). The system is designed to be flexible enough to enable automatic collection and analysis of data on two independent channels, whilst allowing reanalysis of predefined samples stored on disc. Up to four integrators can

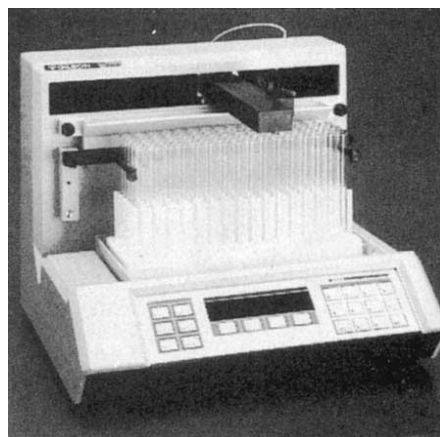


TRIO links up to eight gas chromatographs.

be tied into one data storage unit, thus connecting eight chromatographs. The \$4,300 (US) system may be expanded by adding an additional disc subsystem, giving each TRIO its own floppy disc drive.

Hewlett-Packard has added a flame photometric detector as an option to the HP 5890A gas chromatograph (Reader Service No. 102). The HP 19256A provides selective detection of sulphur and phosphorus compounds by burning them in a hydrogen flame and measuring the emitted light with a photomultiplier tube. It uses a stainless steel burning chamber and a fused-silica transfer line, reducing sample degradation by eliminating contact with hot metal surfaces. The \$5,570 (US) flame photometric detector (\$5,600 when purchased with a HP gas chromatograph) includes adaptors for 0.25 inch glass, 0.125 inch Teflon and 100 to 530 µm fused-silica capillary columns, to allow users to continue using their existing analytical columns and methods.

The Gilson Medical Electronics, Inc. FC 203 fraction collector is designed to be convenient to the chromatographer (Reader Service No. 103). It offers five



The FC 203 can be left alone to do its job.

choices of mode, including time, drop and peak, and key parameters can be edited during the run. The stationary rack system accepts 14 different racks to accommodate vessels ranging in size from 96-well microtitre plates to 13 × 100 mm tubes. The \$1,395 (US) FC 203 interfaces to allow collector control of peripherals or computer control of the collector, and the analog input may be used for data acquisition with Gilson HPLC software.

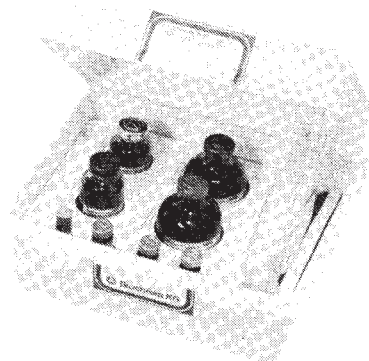
Fast, in-laboratory preparative HPLC column packing is now possible with Sep-Tech's new Macrobores A/E column (Reader Service No. 104). According to

ADVERTISEMENT

A new Principle revolutionizes Immuno- Diagnostics

HIV-ELISA (HIV/HTLV-III/LAV)

- **SYNTHETIC PEPTIDES**
(p18, p24, gp41, gp120)
adsorbed to microtiter
plates
- **no cellular impurities**
- **extra-ordinary high antigen
density**
- **excellent stability and
reproducibility**



Serologically active regions of HIV structural proteins were calculated by means of computer programs. These regions were then synthesized from amino acids on solid phase peptide synthesizers. The serologically most active peptides were chosen from different candidates of all important structural proteins. They were mixed and adsorbed to microtiter plates.

Biochrom KG:
Products for the cell culture
Synthetic Peptides
Diagnostics Reader Service No.18

Biochrom KG

Leonorenstr. 2-6 · D-1000 Berlin (West) 46
Tel. 0 30/7 71 80 47 · Telex 185 821 bio d

SepTech, researchers can pack their own columns in 30 minutes, using the annular expansion system that exerts both radial and axial forces within the column to ensure a uniform and tightly packed media bed. The Macrobore A/E columns are available in volumes ranging from 175 g to 35 kg, and can accommodate packing media with particle sizes between 10 μ m to 100 μ m. All of the columns are dry-packed, eliminating the need for expensive slurry packing equipment. The costs of the columns are dependent upon size, but as a special introduction, the 3 $\times$ 10 in. column is available for \$2,500 (US) through the end of March 1987.

Precise **temperature control of columns** in size exclusion chromatography, amino acid analysis, and high speed separations allows for improved reproducibility and



Warms or cools columns for optimal results.

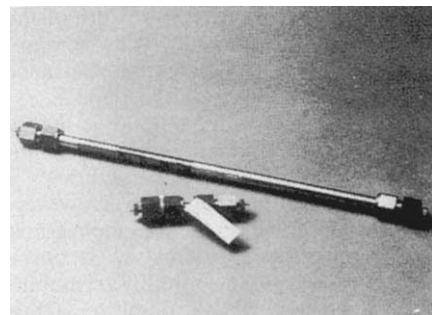
enhanced separations. Waters Chromatography, a division of Millipore Corp., has recently introduced its RCM-100 column heater, capable of holding two 3.9 mm $\times$ 15 cm or 7.8 mm $\times$ 15 cm stainless steel analytical or semi-preparative HPLC columns (Reader Service No. 105). The \$2,000 (US) column heater has a temperature range of 15°C to 70°C, controlled by a Waters temperature control module. This allows the separation of temperature sensitive samples, such as biologicals.

Dionex Ltd claims to have combined the **best of both gas chromatography and liquid chromatography** with its Model 501 Capillary Supercritical Fluid Chromatograph (Reader Service No. 106). Supercritical

fluid chromatography takes advantage of the liquid/vapour equilibrium of a substance, where beyond its critical point it cannot be condensed, whatever the atmospheric pressure. At increasing pressures, the fluid's density, and therefore its solubility, increases, allowing it to behave like a liquid while retaining gas-like diffusivity and viscosity. These traits make supercritical fluids excellent choices as mobile phases for the chromatography of high molecular weight or thermally labile compounds. The £40,000 (UK) Dionex Model 501 consists of separate pump and oven modules, and includes either an IBM or Hewlett Packard PC and associated software.

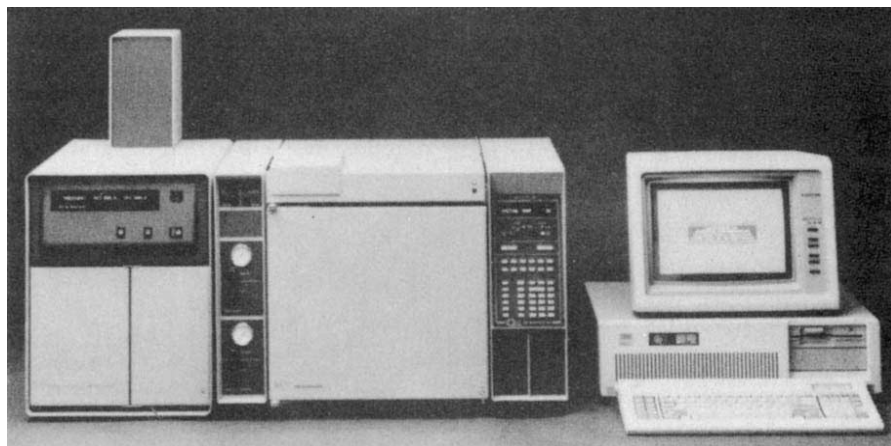
Columns count

A new series of **high resolution size exclusion columns** are available from Beckman Instruments, Inc., expressly for the separation of biopolymers (Reader Service No. 107). The Spherogel-TSK PWHR columns have a particle size of 6–13 μ m to maximize the resolution of proteins, polypeptides, and polynucleotides. Beckman says the columns are stable in the pH range of 2 to 12 with excellent compatibility with acidic, neutral, and alkaline buffers for easy column clean-up. The series includes a mixed bed column for the separation of samples with a wide pH



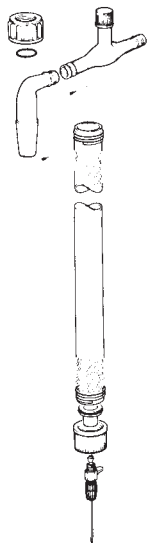
The Beckman Spherogel-TSK PWHR offers ideal separation conditions for biopolymers.

range that eliminates the need to connect columns in a series, shortening analysis time. The 7.8 $\times$ 300 mm columns, available with a variety of packing materials, are priced at \$620 (US).



Supercritical fluid chromatography makes the best of both worlds.

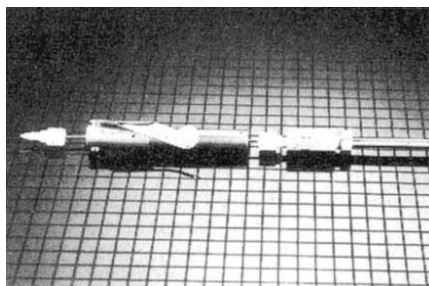
Flash chromatography is ideal for isolating organic compounds from crude reaction products, and Technicol has recently developed a new range of columns and packing materials to aid the flash chromatographer (Reader Service No. 108). Technicol's columns come in a wide



Resolves organics in a flash.

range of sizes, from 18 mm to 100 mm i.d. (15 cm in length), and incorporate plastic sleeves for greater user safety at operating pressures of 5–10 p.s.i. with compressed air or nitrogen. Both silica and bonded silica packing materials are available, with particle sizes ranging from 32 μm to 63 μm and a choice of pore diameters of 60Å, 100Å or 250Å. Technicol claims its flash chromatography products can separate compounds with an Rf difference of just 0.1 within 10 to 15 minutes. The prices of the columns range from £50 to £200 (UK), and the packing materials sell for £16 to £200 (UK).

Two new **guard columns** by Chrompack are now on the market that can be used with any type of analytical column and can be exchanged within seconds, without

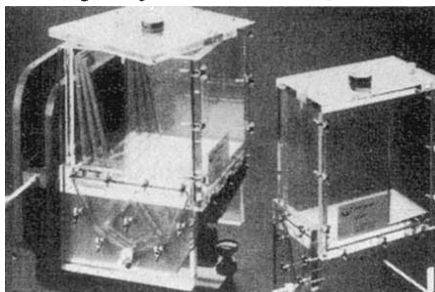


Guards against column damage.

tools (Reader Service No. 109). Chrompack says the columns have far less efficiency loss than conventional guard columns: the high efficiency 5 cm column gives no measureable efficiency loss, while the high capacity 10 cm column (for dirty samples) gives an 8 per cent loss of efficiency. The columns are available with all silica-based packings, and prices range from £40 (UK) for a set of five 10 cm columns to £114 (UK) for five 5 cm columns.

Electrophoretics

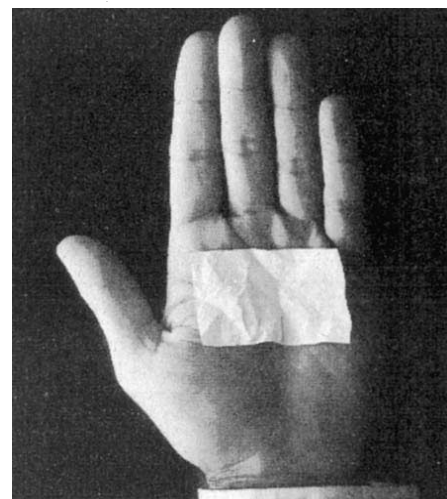
The **gel casters** available from Health Products, Inc. cast 12 or 25 gradient or non-gradient SDS slab gels for the performance of the second dimension in two-dimensional electrophoresis (Reader Service No. 110). Multiple casting of gels ensures the constancy of the gradient of pore size between gels, increasing the reproducibility and standardization of results. The units are preassembled with 7 × 7 inch glass plates, eliminating the need



Makes gradient gels a batch at a time

for tape, grease, or clamps to space the gels. The gel caster base provides support and aids in the uniform delivery of acrylamide to assure the accuracy of the gel pore in gradient gels. The Gel Caster-12 and the Gel Caster-25 cost \$540 and \$595 (US), respectively. The one litre gradient former is \$340 (US), and the two litre gradient former is \$370 (US).

By reinforcing traditional nitrocellulose blotting membranes with a polyester support matrix, Micron Separations Inc. has come up with a **durable membrane** that won't shatter into bits (Reader Service No. 111). The NitroPlus 2000 membranes



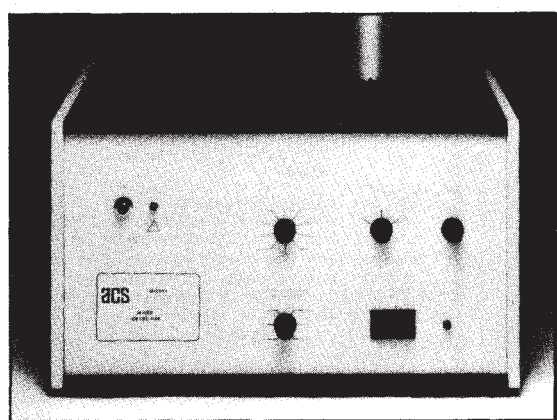
Even after being crumpled, the NitroPlus 2000 refuses to break up.

are purported to have a test strength 500 times that of those composed of unsupported nitrocellulose, enabling them to stand up to reprobing and multiple washing steps. Data loss from membrane breakage in Southern, Northern, and Western blotting is thus avoided. Micron Separations, Inc. is so proud of NitroPlus 2000, it is offering a free sample of the product, and a guarantee that the membrane will not break under standard transfer procedures. The \$40–170 (US) mem-

ADVERTISEMENT

Evaporative (Mass) Detector for GPC and HPLC

- Unique Light Scattering Detection Concept
- Ideal for GPC detection of polymers, carbohydrates, lipids etc.
- CAN be used with Gradient Elution
- NO temperature effects
- Highly sensitive (to ng levels)



Send now for full details

Applied Chromatography Systems Ltd

The Arsenal, Heapy Street
Macclesfield
Cheshire SK11 7JB
Tel 0625 34575
Tx 668005

ACS

PRODUCT REVIEW

branes come in two pore sizes, 0.2 μm and 0.45 μm , and are offered in rolls, discs, rectangles and squares.

The Zeineh soft laser scanning densitometers sold by Biomed Instruments Inc. have wide applications in the analysis of a myriad of separation techniques (Reader Service No. 112). The densitometers may



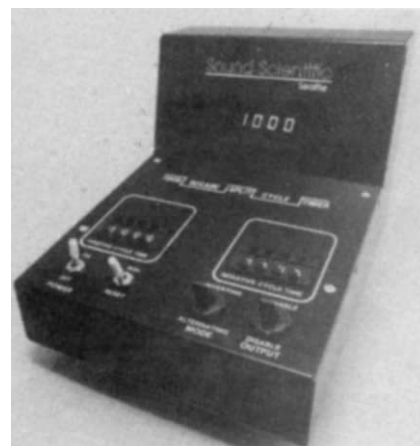
The SLR-2D-1D is good for reading gels and more

be used with all types of electrophoresis gel plates, tubes or capillaries, TLC, autoradiograms, and Western and Southern blots, with a variety of stains from silver

and Coomassie blue to fluorescent stains. Biomed's SLR-2D-1D densitometer offers 3 μm resolution and two-dimensional graphics, with a sensitivity of 5 pg silver stain. Biomed claims the scan speed ranges from 0.001 to 22.6 mm s^{-1} , so a 20 cm $\times$ 20 cm sample can be scanned in 18 minutes. The \$14,950 (US) SLR-2D-1D densitometer has an optional software package for data acquisition and analysis, compatible with the IBM PC and Apple computers.

Stratech Scientific Ltd is now distributing a Savant slab gel dryer, especially designed for laboratories running long sequencing gels for separating and identifying nucleic acid fragments (Reader Service No. 113). The £1,100 (UK) gel dryer features a 40 cm $\times$ 50 cm drying surface, variable heat control, an adjustable timer, two gel supports and a silicone rubber vacuum seal. Options include a mechanically refrigerated trap and a vacuum pump.

The Model H intervalometer developed by Sound Scientific is ideal for use in field inversion gel electrophoresis, according to



Really switches up pulsed field gels.

the manufacturer (Reader Service No. 114). The \$495 (US) Model H can control switching intervals independently from 1 to 9,999 seconds, and provides ultra-low frequency pulses of up to 110 VAC at 3A or 500 VDC at 500 mA.

Brownlee Labs has begun circulation of a quarterly newsletter devoted to problems and technical advice on chromatography (Reader Service No. 115). The first issue of *Chromatography* details the intricacies of reversed phase HPLC, including recommendations to help solve the dilemma of which column to use. Available free of charge to the scientific community.

Coming in Product Review...

● **Peptides and oligonucleotides.** An interesting array of the new equipment and reagents on exhibit at the DNA Winter Symposium in Miami, Florida, will be featured on 5 February, along with an article on the Bionet electronic computer network.

● **Lasers.** Showcased on 19 February will be the latest in the development and application of laser technology in the laboratory.

● **Microbiology.** In conjunction with the meeting of the American Society for Microbiology in Atlanta, Georgia, the 26 February issue will outline products of use to researchers working with bacteria and viruses.

● **Spectroscopy.** A variety of things of use to spectroscopists will be reviewed on 5 March, with emphasis on those products exhibited at the Pittsburgh Conference in Atlantic City, New Jersey. □

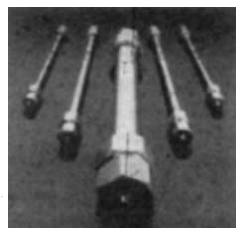
These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

ADVERTISEMENTS

EXCEL HPLC COLUMNS

3 μm ODS over 130,000 plates/metre
5 μm ODS over 80,000 plates/metre

NEW



- UNBEATABLE EFFICIENCY
- SPHERICAL 3 μm AND 5 μm ODS BONDED SILICAS
- OPTIMUM BATCH SELECTION
- UNIQUE PACKING PROCESS
- SPECIALLY POLISHED APOLLO TUBING
- TRIPPLICATE PERFORMANCE TESTING

FOR FURTHER DETAILS CALL HICHROM

TEL: READING (0734) 303660

HICHROM LTD, 6 Chiltern Enterprise Centre, Station Road, Theale, Reading, Berkshire, RG7 4AA. Tlx 848210-INFOS-G-HICHROM

Reader Service No.24

WESTERN BLOTS IN MINUTES

Our new GENIE electrophoretic blotter incorporates space-age plate electrodes for a uniform, high-intensity field. It is easy to load, and uses less than a liter of buffer, yet can blot gels up to 15 by 17 cm. in as little as ten minutes.

only \$190



Idea Scientific Company
P.O. Box 2078
Corvallis, Oregon 97339
(503) 758-0999

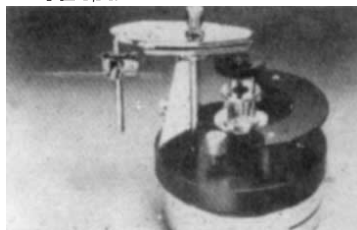
ASM Booth 1715

Reader Service No.14

SINGER INSTRUMENTS

MICROMANIPULATORS & ACCESSORIES

YEAST? TETRAD ANALYSIS?



Singer MK III Micromanipulator and Precision Elevator.

Write direct to:

Singer Instrument Co. Ltd.

Treborough Lodge
Roadwater, Watchet
Somerset TA23 0QL
England. Tel. (0984) 40226
Telex: 337492 Comcab G

Reader Service No.50

LIPOSOMES

REPRODUCIBLE - HOMOGENEOUS - EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

BIANORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.47